Wavelet analysis of temporal data

David Alexander Goodwin

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

Department of Statistics

February 2008

Abstract

This thesis considers the application of wavelets to problems involving multiple series of temporal data. Wavelets have proven to be highly effective at extracting frequency information from data. Their multi-scale nature enables the efficient description of both transient and long-term signals. Furthermore, only a small number of wavelet coefficients are needed to describe complicated signals and the wavelet transform is computationally efficient.

In problems where frequency properties are known to be important, it is proposed that a modelling approach which attempts to explain the response in terms of a multi-scale wavelet representation of the explanatory series will be an improvement on standard regression techniques. The problem with classical regression is that differing frequency characteristics are not exploited and make the estimates of the model parameters less stable. The proposed modelling method is presented with application to examples from seismology and tomography.

In the first part of the thesis, we investigate the use of the non-decimated wavelet transform in the modelling of data produced from a simulated seismology study. The fact that elastic waves travel with different velocities in different rock types is exploited and wavelet models are proposed to avoid the complication of predictions being unstable to small changes in the input data, that is an inverse problem.

The second part of the thesis uses the non-decimated wavelet transform to model electrical tomographic data, with the aim of process control. In industrial applications of electrical tomography, multiple voltages are recorded between electrodes attached to the boundary of, for example, a pipe. The usual first step of the analysis is then to reconstruct the conductivity distribution within the pipe. The most commonly used approaches again lead to inverse problems, and wavelet models are again used here to overcome this difficulty. We conclude by developing the non-decimated multi-wavelet transform for use in the

modelling processes and investigate the improvements over scalar wavelets.

I would like to thank my supervisors, Doctors Robert Aykroyd and Stuart Barber, for their support and enthusiasm over the last four years. Many of the ideas in the thesis originally came from my discussions with them. I would also like to thank Doctor Paul Baxter for being my internal assessor and providing useful suggestions during my annual reviews. I am grateful to the EPSRC for providing funding to support my research. Finally, I would like to thank Cat, my wife, as it would not have been written without her support and patience.

If we knew what we were doing, it wouldn’t be called research, would it?

— *Albert Einstein*

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation

The analysis of temporal data can be complicated when the data have properties that make the application of the usual time series methods such as ARIMA models introduced by Box and Jenkins (1976) and spectral analysis Priestley (1981) difficult. An example of this would be a series lacking stationarity in terms of its mean value or autocovariance structure over time. Seasonal time series, such as rainfall, are examples of non-stationary series.

When a process consists of more than one observed time series, the methods for modelling and investigating the properties get increasingly difficult as the number of series increases.

The idea of a multiscale approach will also prove useful when analysing temporal data which contains several features, each occurring over different time scales. The example of analysing environmental data illustrates how this approach may be of benefit. Whilst climate change might be the cause of temperature increases over decades, season changes will bring about temperature change throughout the year and consideration of the daily weather provides a picture at a small scale. The wavelet transform of such a dataset would

decompose the series into several sub-series, each offering information at a different time scale.

1.2 Outline of thesis

The thesis begins with an introduction to wavelet methods in Chapter 2. This includes an overview of the wavelet transform and the performance of some noise estimation methods is explored.

The new contributions made in this thesis are in developing methods for modelling situations where signals have changing frequency properties at given points in time, using wavelet decompositions to transform the underlying signals may allow the frequency properties to be incorporated into the modelling paradigm. By incorporating what is to be called a 'transfer function', it has been possible to effectively encode the frequency information.

This thesis considers several situations where the underlying process produces data with such frequency properties and proposes a framework in each setting to best allow the properties to be exploited.

Chapter 3 discusses a method of transforming seismic data that allows the varying frequency properties to be exploited. These are then incorporated into a linear model to predict acoustic impedance.

Chapter 4 applies wavelet methods to tomographic data modelling. The question of how to incorporate multiple observations into the modelling framework is investigation.

An extension to wavelet theory, known as multiwavelets is explored in Chapter 5. Here an alternative method of incorporating many series into a modelling framework is explored in the seismic and tomographic settings. An overview of the mathematical background to multiwavelets can be found in Keinert (2004).

The proposed methods could have possible applications in other data settings which exhibit similar properties. Examples could include the modelling of customer behaviour modelling in financial services, where higher frequency activity could indicate a change in behaviour indicative of either cross selling opportunities or customers leaving. Environmental data, such as wind speeds or rainfall may also exhibit changing frequencies due to the different weather systems.

1.3 New developments

The thesis develops new methodology to model series with both frequency and time characteristics through a combination of wavelet methods and existing statistical modelling techniques such as linear and logistic regression. The models are built using processed wavelet coefficients and the proposed method of processing, using a transfer function (described in Section 3.4), is also a new concept.

Chapter 2

Literature Review

2.1 Introduction

Wavelets are localised basis functions, with some special mathematical properties which will be mentioned later in section 2.4. An analogy for how wavelets work is to think of a camera lens that allows you to take broad landscape pictures as well as zoom in on microscopic detail that can't normally be seen by the human eye.

Many applications of wavelets have been developed over recent years and now the applications range from Biomedical imaging Van der Ville et al. (2006) and microarray analysis X.H.Wang et al. (2003) to data fusion Fryzlewicz et al. (2007).

A usual starting point to explain how wavelets work is to start with the ideas of Fourier Theory, which represent functions in terms of a series of sine and cosine functions (having infinite support).

2.2 Fourier Analysis

It is known from Fourier theory that a signal can be expressed as the sum of a series of sines and cosines, known as a Fourier expansion Körner (1996). However, these methods fail to provide efficient representations for certain types of functions which have discontinuities.

Definition 2.2.1 *A Fourier series is the representation of a function, f, in terms of a sum of sine and cosine functions*

$$f(x) = \frac{a_0}{2\pi} + \frac{1}{\pi}\sum_{n=1}^{\infty}(a_n \cos(nx) + bn \sin(nx))$$

where

$$\begin{aligned} a_n &= \int_0^{2\pi} f(x)\cos(nx)dx \\ b_n &= \int_0^{2\pi} f(x)\sin(nx)dx \end{aligned}$$

This representation is possible since $\{1, \cos(nx), \sin(nx)\}$ constitutes an orthonormal basis of $\mathbb{L}^2([0, 2\pi])$.

Sine and cosines functions (sinusoids) are localised in frequency but not in time, thus a large number of terms are required to represent a function with a discontinuity in terms of a Fourier series.

Definition 2.2.2 *Let $< f, g >= \int f(x)\overline{g(x)}dx$ denote the inner product of the functions f and g. The Fourier transform of a function $f \in \mathbb{L}^1(\mathbb{R})$ is defined by*

$$\begin{aligned}\hat{f}(\omega) &= \mathcal{F}[f(\omega)] \\ &= < f(x), e^{i\omega x} > \\ &= \int_{\mathbb{R}} f(x)e^{-i\omega x}dx\end{aligned}$$

If $\hat{f} \in \mathbb{L}^1(\mathbb{R})$ is the Fourier transform of $f \in \mathbb{L}^1(\mathbb{R})$ then

$$f(x) = \mathcal{F}^{-1}[\hat{f}(\omega)] = \frac{1}{2\pi}\int_{\mathbb{R}} \hat{f}(\omega)e^{i\omega x}d\omega$$

is the inverse Fourier transform.

The disadvantage of a Fourier expansion is that it has only frequency resolution and no time resolution. This means that although it is possible to determine all the frequencies present in a signal, it is not possible to know when they are present. To overcome this problem wavelets can be used. In order to understand how a wavelet decomposition can be constructed, the concept of a multiresolution analysis will be introduced.

2.3 Multiresolution Analysis

A multiresolution analysis (MRA) provides a framework for examining functions at different scales. A multiresolution analysis can be defined following Mallat (1989).

Definition 2.3.1 *A multiresolution analysis is a chain of nested closed subspaces,* $\{V_j, j \in \mathbb{Z}\}$ *satisfying the following conditions:*

1. *The spaces have trivial intersection:*

$$\bigcap_{j\in\mathbb{Z}} V_j = \{0\}.$$

2. *The union is dense in $L^2(\mathbb{R})$:*

$$\bigcup_{j\in\mathbb{Z}} V_j = L^2(\mathbb{R}).$$

3. *The following scale relations exist:*

$$f(x) \in V_j \quad \Leftrightarrow \quad f(2x) \in V_{j+1}. \tag{2.1}$$

$$f(x) \in V_0 \quad \Leftrightarrow \quad f(x-k) \in V_0,\ k \in \mathbb{Z}. \tag{2.2}$$

4. *There exists a function $\phi(x) \in V_0$ such that the sequence $\{\phi(x-k),\ k \in \mathbb{Z}\}$ is an orthonormal basis of V_0.*

The conditions given in equations 2.1 and 2.2 imply that $\{\phi_{jk},\ k \in \mathbb{Z}\}$ is an orthonormal basis of V_j. Since $V_0 \subset V_1$, the function $\phi(x) \in V_0$ can be represented as a linear combination of the functions from V_1 and so

$$\phi(x) = \sum_{k\in\mathbb{Z}} h_k \sqrt{2}\phi(2x-k),$$

for some coefficients h_k, $k \in \mathbb{Z}$, which are often referred to as a low pass filter. The function $\phi(x)$ is called the *father wavelet* or *scaling function.*

For each MRA, it is also possible to define a *mother wavelet*, $\psi(x)$, which will explain the detail at each level j. Consider the detail space W_j to be the orthogonal complement of V_j in V_{j+1}, so that

$$V_{j+1} = V_j \oplus W_j.$$

Then $\{\psi(x-k),\ k \in \mathbb{Z}\}$ forms an orthonormal basis for W_0, orthogonal to all functions in V_j. Since $\psi(x) \in V_1$, the function $\psi(x)$ can similarly be represented as a linear combination of the functions from V_1.

$$\psi(x) = \sum_{k\in\mathbb{Z}} g_k \sqrt{2}\phi(2x-k),$$

where the coefficients g_k, $k \in \mathbb{Z}$ are known as the high pass filter. As a consequence of the MRA conditions, $f(x) \in W_0 \Leftrightarrow f(2_j x) \in W_j$ and so $\{\psi_{jk}(x)\}$ is an orthonormal basis for W_j and so

$$\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k),\ j, k \in \mathbb{Z}.$$

2.4 What is a wavelet?

A wavelet is a small localised wave, designed to have attractive properties not enjoyed by "big waves" such as short support which is not exhibited by the sinusoids.

Wavelets are basis functions which are able to represent a signal in the time and frequency domain at the same time. They can be used to approximate an underlying trace or signal, similar to Fourier transforms. The advantages of wavelets are that they are localised in frequency and time and so can handle a wider range of signals than Fourier analysis. A disadvantage of wavelets is that the transform obtained only has representations of the data at a discrete number of resolution levels, each resolution level having a representation at approximately twice the frequency of the previous level.

A wavelet basis can be formed by translating and dilating a function ψ, called the mother wavelet.

A mother wavelet may be defined as in Meyer (1992).

Definition 2.4.1 *Let $m \in \mathbb{N}$. Then for $x \in \mathbb{R}$, a function $\psi(x)$ is called a mother wavelet of order m if the following properties hold.*

1. *If $m = 0, \psi(x) \in L^\infty(\mathbb{R})$. If $m \geq 1$, then $\psi(x)$ and all its derivatives up to order m belong to $\mathbb{L}^\infty(\mathbb{R})$.*

2. *$\psi(x)$ and all its derivatives up to order m decrease rapidly as $x \to \pm\infty$.*

3. *For each* $k \in \{0, \ldots, m\}$,
$$\int_{-\infty}^{\infty} x^k \psi(x) dx = 0$$

4. *The collection* $\{\psi_{j,k}\}_{j,k \in \mathbb{Z}}$ *forms an orthonormal basis of* $\mathbb{L}^2(\mathbb{R})$*, the* $\psi_{j,k}$ *being constructed from the mother wavelet using*
$$\psi_{j,k}(x) = 2^{\frac{j}{2}} \psi(2^j x - k).$$

2.5 Examples of wavelet bases

2.5.1 The Haar basis

Probably the simplest wavelet basis is the Haar basis, constructed from following the scaling function and mother wavelet. A picture of the Haar basis is shown in Figure 2.1.

$$\psi(x) = \begin{cases} 1 & x \in (0,1) \\ 0 & x \notin (0,1) \end{cases}, \ \phi(x) = \begin{cases} 1 & x \in [0, 0.5) \\ -1 & x \in [0.5, 1) \\ 0 & \text{otherwise} \end{cases}$$

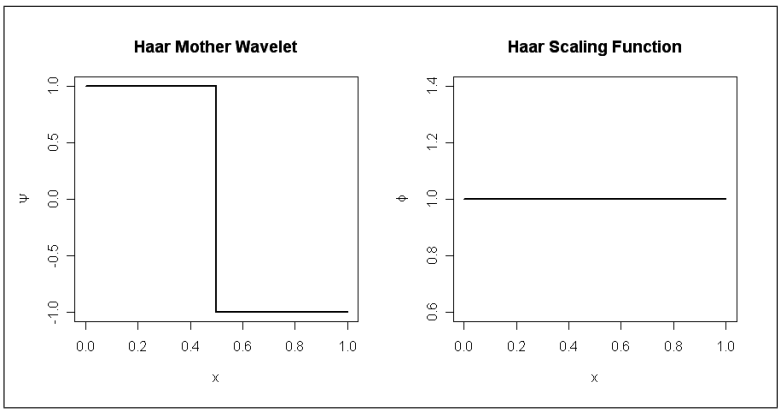


Figure 2.1: Picture of the Haar mother wavelet and scaling function.

2.5.2 The Meyer wavelet

This is thought about in terms of the frequency domain.

$$\widehat{\psi}(x) = \begin{cases} (2\pi)^{-\frac{1}{2}} & |x| \leq \frac{2\pi}{3} \\ (2\pi)^{-\frac{1}{2}} \cos\left[\frac{\pi}{2}\nu\left(\frac{3}{2\pi}|x| - 1\right)\right] & \frac{2\pi}{3} \leq |w| \leq \frac{4\pi}{3} \\ 0 & \text{otherwise} \end{cases}$$

where ν is a smooth function satisfying:

$$\nu(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x \geq 1 \end{cases}$$

and $\nu(x) + \nu(1 - x) = 1$. The wavelet is of the form:

$$\widehat{\psi}(x) = \sqrt{2\pi} e^{\frac{ix}{2}} [\widehat{\phi}(x + 2\pi) + \widehat{\phi}(x - 2\pi)] \widehat{\phi}\left(\frac{x}{2}\right)$$

2.5.3 Daubechies wavelets

Daubechies (1992) introduced two families of compactly supported wavelets, with different possible degrees of smoothness. These are known as Daubechies *extremal phase* and *least asymmetric* wavelets. A useful property of compactly supported wavelets is that the associated filters $\mathcal{H}$ and $\mathcal{G}$, which will be explained in Section 2.6, have a finite number of coefficients. More details of the construction of these wavelets can be found in Daubechies (1992) and some examples of Daubechies compactly supported mother wavelets are shown in Figure 2.2.

Further details of different wavelet bases can be found in Vidakovic (1999) and Percival and Walden (2000).

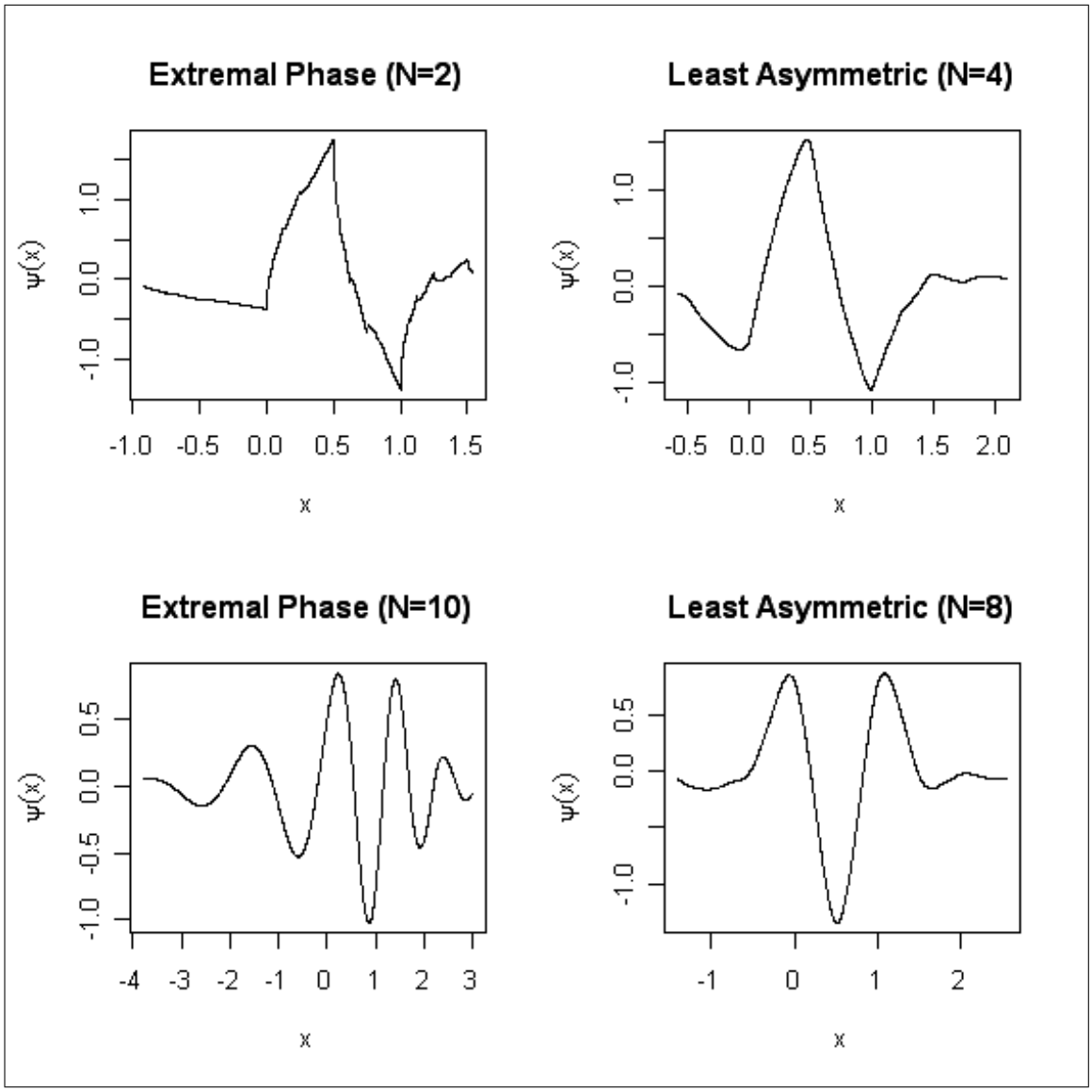


Figure 2.2: Examples of Daubechies compactly supported mother wavelets.

2.6 The discrete wavelet transform

The discrete wavelet transform (DWT), proposed by Mallat (1989), is an efficient algorithm for calculating the wavelet coefficients of a discrete series. The idea is to filter the series, using the high and low pass filters associated with the wavelet basis to obtain the wavelet coefficients.

The transform can be explained in terms of a low pass filter $\mathcal{H} = \{h_k\}$ and a high pass filter $\mathcal{G} = \{g_k\}$, where the h_k and g_k are the coefficients of the filters referred to in Section 2.3.

Consider a function f observed at $N = 2^J$ equally spaced time points $\{t_i, i = 0, \ldots N - 1\}$. Let $c_{J,i} = f(t_i)$ for $i = 0, \ldots, N - 1$. The discrete wavelet transform of the series can be obtained using the relations:

$$c_{j-1,i} = \sum_n h_{n-2i} c_{j,n} \tag{2.3}$$

$$d_{j-1,i} = \sum_n g_{n-2i} c_{j,n} \tag{2.4}$$

The resulting wavelet transform is the collection of the detail coefficients at each level together with the smooth or father coefficient at the zero level. This is an orthogonally transformed representation of the original series of length N. The $2i$ term in equations 2.3 and 2.4 is an alternative way of representing the decimation step of the DWT. It is equivalent to filtering using $\mathcal{H}$ and $\mathcal{G}$ and then applying even dyadic decimation (selecting every even observation from the series).

An alternative way to formulate the DWT is to construct an orthogonal matrix $\mathbf{W}$ associated with the particular wavelet basis being used. The DWT can then be defined as the matrix multiplication of this orthogonal matrix with a vector of observation points, $\mathbf{x}$. That is,

$$\mathbf{d} = \mathbf{W}\mathbf{x},$$

where $\mathbf{d}$ is a vector comprising of both the discrete mother and father wavelet coefficients.

2.6.1 Example of the DWT

Suppose we have a four element sequence

$$(c_{2,0}, c_{2,1}, c_{2,2}, c_{2,3}) = (4, 12, 6, 24)$$

and we want to find the wavelet decomposition using the Haar basis. The filters for the Haar basis are

$$\mathcal{H} = (h_0, h_1) = (1/\sqrt{2}, 1/\sqrt{2}),\ \mathcal{G} = (g_0, g_1) = (1/\sqrt{2}, -1/\sqrt{2}).$$

Using Equations 2.3 and 2.4, we obtain

$$\begin{aligned} c_{1,0} &= 1/\sqrt{2} \times 4 + 1/\sqrt{2} \times 12 = 8\sqrt{2} \\ c_{1,1} &= 1/\sqrt{2} \times 6 + 1/\sqrt{2} \times 24 = 15\sqrt{2} \\ d_{1,0} &= 1/\sqrt{2} \times 4 - 1/\sqrt{2} \times 12 = -4\sqrt{2} \\ d_{1,1} &= 1/\sqrt{2} \times 6 - 1/\sqrt{2} \times 24 = -9\sqrt{2} \end{aligned}$$

At the next resolution level, the coefficients are

$$\begin{aligned} c_{0,0} &= 1/\sqrt{2} \times 8\sqrt{2} + 1/\sqrt{2} \times 15\sqrt{2} = 23 \\ d_{0,0} &= 1/\sqrt{2} \times 8\sqrt{2} - 1/\sqrt{2} \times 15\sqrt{2} = -7. \end{aligned}$$

It is possible to reconstruct the original series from the coefficients $(c_{0,0}, d_{0,0}, d_{1,0}, d_{1,1})$ using the inverse DWT.

2.6.2 Boundary conditions and wavelets on the interval

The Haar wavelet has very short filters, but for other bases, a problem may occur when the filters are longer than the input series. Several solutions exist to this problem: the series

could be considered to repeat periodically; padding with zeroes could increase the length of the series; polynomial extrapolation could be used. Nason and Silverman (1994) give more details on the options available.

Meyer (1991) showed how an orthonormal family of wavelets in $[0, 1]$ can be constructed from any compactly supported wavelet basis of Daubechies (1988). The resulting basis has the same number of vanishing moments and the same regularity as the original mother wavelet. An orthonormal basis for $\mathbb{L}^2([0, 1])$ can be formed from these functions, together with the father wavelets as the coarsest scale. Constructing wavelets on the interval in this way has the disadvantage that in the explicit construction involves diagonalisation of a matrix and becomes ill conditioned (Cohen et al., 1993).

Cohen et al. (1993) proposed a modification in which the resulting wavelets are derived from the minimal compactly supported wavelet for $\mathbb{R}$. More recently, Melkemi and Mokhtari (2007) have developed a general method to construct wavelet bases on an interval with arbitrary support.

2.7 Inverse wavelet transform

When the wavelet transform is thought of as a matrix multiplication, it is easy to see that the orthogonal matrix $\mathbf{W}$ can be inverted and so the inverse transform can be found. Computationally, the inverse transform takes the coarsest level father and mother coefficients and uses them to reconstruct the next finer level. Mathematically this can be written as

$$c_{j,k} = \sum_l h_{k-2l} c_{j-1,l} + \sum_l g_{k-2l} d_{j-1,l}.$$

The reconstruction can then be achieved by iterating this process and climbing the resolution levels back to the original data.

2.8 Non-decimated wavelet transform

The first application of the non-decimated (or stationary) wavelet transform (NDWT) to statistical problems was by Nason and Silverman (1995). To compute the NDWT, appropriate high and low pass filters are applied to the data at each level to produce two sequences at the next level. No decimation occurs in this transform, so the two sequences have the same length as the original sequence. Instead, the filters are modified at each level, by padding out with zeroes.

2.8.1 Setup

Define Operators

1. Let $\mathcal{Z}$ be an operator which alternates a given sequence with zeroes, so that

$$\forall j \in \mathbb{Z}, (\mathcal{Z}x)_{2j} = x_j \text{ and } (\mathcal{Z}x)_{2j+1} = 0$$

2. Let $\mathcal{S}$ be the shift operator defined by

$$(\mathcal{S}x)_j = x_{j+1}$$

3. Let $\mathcal{D}_0$ be the binary decimation operator defined by

$$(\mathcal{D}_0 x)_j = x_{2j}$$

Define filters

Define $\mathcal{H}^{[r]}$ and $\mathcal{G}^{[r]}$ to have weights $\mathcal{Z}^r h$ and $\mathcal{Z}^r g$ respectively. Thus the filter $\mathcal{H}^{[r]}$ has the weights $h^{[r]}_{2^r j} = h_j$ and $h^{[r]}_k = 0$ if k is not a multiple of 2^r.

So $\mathcal{H}^{[r]}$ is obtained by inserting a zero between every adjacent pair of elements of the filter $\mathcal{H}^{[r-1]}$, and similarly for $\mathcal{G}^{[r]}$.

$$\Rightarrow \mathcal{H}^{[r]} \text{ and } \mathcal{G}^{[r]} \text{ commute with } \mathcal{S}$$

$$\Rightarrow \mathcal{D}_0^r \mathcal{H}^{[r]} = \mathcal{H}\mathcal{D}_0^r \text{ and } \mathcal{D}_0^r \mathcal{G}^{[r]} = \mathcal{G}\mathcal{D}_0^r$$

Defining the NDWT

Let a^J be a sequence. For $j = J, J-1, \ldots, 1$ define $a^{j-1} = \mathcal{H}^{[J-j]} a^j$ and $b^{j-1} = \mathcal{G}^{[J-j]} a^j$.

If a^J is of length 2^J, then all the vectors a^j and b^j will be of the same length, rather than becoming shorter as j decreases as in the standard DWT. This means that to find $b^{J-1}, b^{J-2}, \ldots, b^0, a^0$ will take $O(J2^J)$ operations rather than $O(2^J)$ as in the decimated DWT.

2.8.2 Example of the NDWT

Recalling the example from Section 2.6.1, we now find the non-decimated version of the wavelet transform for this series. To allow the effect on the filters to be seen, dot product notation is used.

The scaling coefficients $\{c_{j,k}\}$ are found using periodic boundary conditions from the following relationships.

$$\begin{aligned} c_{1,0} &= (h_0, h_1) \cdot (c_{2,0}, c_{2,1}) \\ c_{1,1} &= (h_0, h_1) \cdot (c_{2,1}, c_{2,2}) \\ c_{1,2} &= (h_0, h_1) \cdot (c_{2,2}, c_{2,3}) \\ c_{1,3} &= (h_0, h_1) \cdot (c_{2,2}, c_{2,0}) \end{aligned}$$

The mother wavelet coefficients can be found similarly using the high pass filter, $\mathcal{G} = (g_0, g_1)$, instead of the low pass.

At the next level, the filters need to padded with alternate zeros before calculating the coefficients, and so the next level can be calculated as follows.

$$\begin{aligned} c_{0,0} &= (h_0, 0, h_1, 0) \cdot (c_{1,0}, c_{1,1}, c_{1,2}, c_{1,3}) \\ c_{0,1} &= (h_0, 0, h_1, 0) \cdot (c_{1,1}, c_{1,2}, c_{1,3}, c_{1,0}) \\ c_{0,2} &= (h_0, 0, h_1, 0) \cdot (c_{1,2}, c_{1,3}, c_{1,0}, c_{1,1}) \\ c_{0,3} &= (h_0, 0, h_1, 0) \cdot (c_{1,3}, c_{1,0}, c_{1,1}, c_{1,2}). \end{aligned}$$

2.8.3 Numerical example

If we have a sequence similar to the example in Section 2.6.1, the NDWT would be calculated as (using the Haar basis):

$$(c_{2,0}, c_{2,1}, c_{2,2}, c_{2,3}) = (4, 12, 6, 24)$$

.

$$\begin{aligned} c_{1,0} &= (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}) \cdot (4, 12) = 8\sqrt{2} \\ c_{1,1} &= (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}) \cdot (12, 6) = 9\sqrt{2} \\ c_{1,2} &= (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}) \cdot (6, 24) = 15\sqrt{2} \\ c_{1,3} &= (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}) \cdot (24, 4) = 14\sqrt{2} \end{aligned}$$

The next level of coefficients can then be found.

$$\begin{aligned} c_{0,0} &= (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}, 0) \cdot (8\sqrt{2}, 9\sqrt{2}, 15\sqrt{2}, 14\sqrt{2}) = 23 \\ c_{0,1} &= (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}, 0) \cdot (9\sqrt{2}, 15\sqrt{2}, 14\sqrt{2}, 8\sqrt{2}) = 23 \\ c_{0,2} &= (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}, 0) \cdot (15\sqrt{2}, 14\sqrt{2}, 8\sqrt{2}, 9\sqrt{2}) = 23 \\ c_{0,3} &= (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}, 0) \cdot (14\sqrt{2}, 8\sqrt{2}, 9\sqrt{2}, 15\sqrt{2}) = 23. \end{aligned}$$

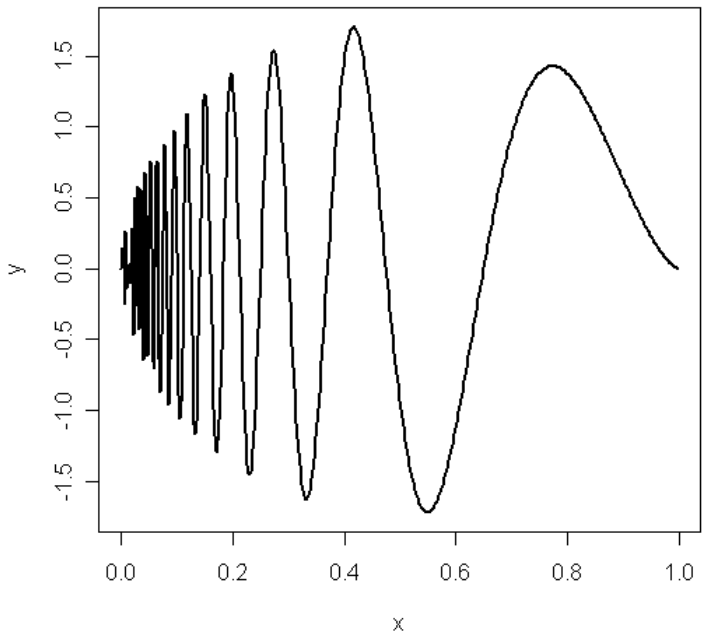
2.8.4 Relation with the DWT

The stationary wavelet transform contains all the coefficients of the decimated wavelet transform. The stationary wavelet transform 'fills in the gaps' between the coefficients in a decimated DWT and so there is no longer any restriction of the localization position to a grid of integers.

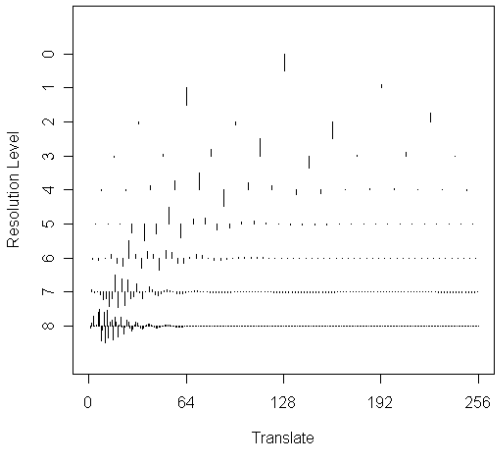
2.8.5 Example of the NDWT as an exploratory method

In this example, the Doppler test signal (Figure 2.3(a)), as discussed in Section 2.9.3 is used. This signal has varying frequency, from high to low along the time axis.

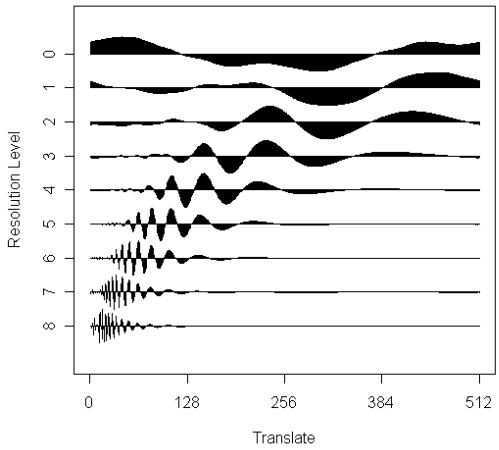
The standard wavelet transform of this signal (Figure 2.3(b)) shows this at the finer resolution levels. The stationary wavelet transform, shown in Figure 2.3(c), also highlights the decreasing frequency. As x increases, the amplitude of the oscillation within the higher frequency levels decreases.



(a) Doppler test signal



(b) Decimated



(c) Non decimated

Figure 2.3: Wavelet decompositions of the Doppler test signal using the Daubechies extremal phase wavelet with 2 vanishing moments.

2.9 Thresholding

Wavelet methods are often used in the non-parametric regression setting, where the problem is of the form:

$$y_i = g(t_i) + \epsilon_i$$

where $t_i = (i-1)/N$, $i = 1, \ldots, N$ and $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$.

How can we recover the function g from the data y_i without assuming any particular parametric structure for g? One answer, introduced by Donoho and Johnstone (1994), is through wavelet thresholding.

The process involved in thresholding is to take the wavelet decomposition of the data $\{y_i\}$. An attempt is then made to identify which of the wavelet coefficients obtained are representing noise and these coefficients are modified according to a thresholding rule in an attempt to remove the noise from the signal. Finally, the inverse wavelet transform is applied to the modified coefficients to obtain an estimate of g.

2.9.1 Hard and soft thresholding

Hard thresholding rules transform all the coefficients regarded as negligible (i.e., those satisfying $|d_{jk}| < \lambda$) to 0, keeping all other coefficients at their original value. The thresholding rule is expressed as:

$$\widehat{d_{jk}} = d^*_{jk}\mathtt{I}(|d^*_{jk}| > \lambda)$$

Soft thresholding rules transform some of the smaller coefficients to 0, and translates the rest towards 0. It is easily expressed by the following equation:

$$\widehat{d_{jk}} = \mathtt{sgn}(d^*_{jk})\,\mathtt{max}(0, |\widehat{d_{jk}}| - \lambda)$$

Hard threshold is a 'keep or kill' procedure and is more intuitively appealing. The alternative, soft thresholding, shrinks coefficients above the threshold in absolute value. While at first sight hard thresholding may seem to be natural, the continuity of soft thresholding has some advantages, one of which is that it makes algorithms mathematically more tractable (Donoho and Johnstone, 1995). Sometimes, pure noise coefficients may pass the hard threshold and appear as 'blips' in the output. Soft thresholding shrinks these false structures. In terms of image de-noising, hard thresholding maintains the sharp edges of the image, whereas soft thresholding tends to blur the image. However, soft thresholding does usually result in reconstructions with lower mean squared error.

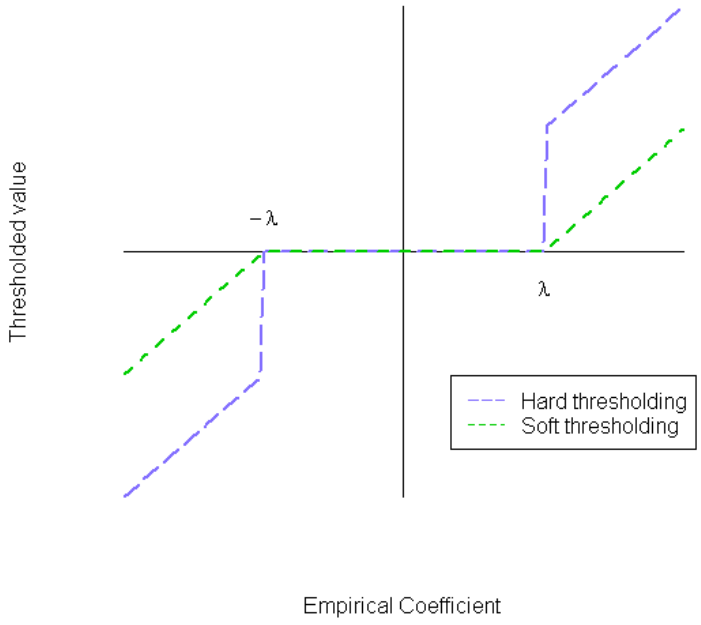


Figure 2.4: Comparison of hard and soft thresholding.

There are alternatives to hard and soft thresholding in the form of localised, context-based thresholding. Ghazel et al. (2005) show that the performance of thresholding can

be improved using such techniques.

2.9.2 How do you choose a threshold?

Many different schemes have been proposed for choosing the threshold, λ. Abramovich et al. (2000) and Vidakovic (1999) give a review of some of these. We give a brief description of some of the standard thresholding rules.

Universal thresholding

The universal threshold was proposed by Donoho and Johnstone (1994). It is given by

$$\lambda_{\text{universal}} = \sigma\sqrt{2\log N},$$

where the value of σ will usually be unknown and therefore estimated from the data.

BayesThresh

The BayesThresh method (Abramovich et al., 1998) tries to estimate the 'large' coefficients and set the others equal to zero. This is achieved by assuming that the wavelet coefficients, d_{jk}, have the following prior distribution:

$$d_{jk} \quad \sim \quad \pi_j N(0, \tau_j^2) \; + \; (1 - \pi_j)\, \delta(0) \tag{2.5}$$

where $0 \leq \pi_j \leq 1$, $\delta(0)$ is a point mass at 0 and the d_{jk}'s are independent.

The model is specified in terms of the hyperparameters τ and π, with these parameters themselves having the following representation in terms of two user selected inputs, α and β:

$$\begin{aligned} \tau_j^2 &= 2^{-\alpha_j} C_1 \\ \pi_j &= \texttt{min}(1, 2^{-\beta_j} C_2) \end{aligned}$$

where C_1 and C_2 are constants chosen empirically from the data.

Abramovich et al. (1998) show that the choices of α and β are equivalent to selecting particular prior distribution to incorporate prior knowledge about the smoothness of the underlying function. Chipman and Wolfson (1999) also discuss the interpretation of α and β.

The implementation of the BayesThresh method (Barber et al., 2002) specifies the prior model in terms of the parameters α and β, calculating the constants C_1 and C_2, from the data.

The prior specification then assumes a non-informative prior for the scaling coefficient $c_{0,0}$, which has the posterior distribution $N(c^*_{0,0}, \sigma^2)$, and is estimated by the observed value $c^*_{0,0}$.

The resulting posterior distribution of d_{jk} given an observed value of d^*_{jk} is independent for each jk and is given by

$$d_{jk}|d^*_{jk} \sim \omega_{jk} N(d^*_{jk} r_j^2, \sigma^2 r_j^2) + (1 - \omega_j)\delta(0)$$

where $\omega_{jk} = 1/(1 - \xi_{jk})$ with

$$\xi_{jk} = \frac{1 - \pi_j}{\pi_j} \frac{\sqrt{\tau_j^2 + \sigma^2}}{\sigma} \exp \frac{-\tau_j^2 d^*_{jk}}{2\sigma^2(\tau_j^2 + \sigma^2)}$$

and $r_j^2 = \tau_j^2/(\sigma^2 + \tau_j^2)$.

SURE

This method of estimating the threshold was proposed by Donoho and Johnstone (1995) and is based upon the minimisation of Stein's unbiased risk estimator. The wavelet coefficients at each level are considered separately as independent multivariate normal

estimation problems. Stein (1981) showed that an unbiased estimator of the risk is

$$\text{SURE}(\lambda_j, d_{jk}) = N - 2\sum_{k=1}^{N} \mathtt{I}(|d_{jk}| \leq \lambda_j) + \sum_{k=1}^{N} \min(|d_{jk}|, \lambda)^2$$

and so the SURE threshold can be written as

$$\lambda_{j,\text{sure}} = \text{argmin}_{0 \leq \lambda \leq \sqrt{2\log N}} \text{SURE}(\lambda_j, d_j k).$$

2.9.3 Estimating the noise standard deviation, σ

There are several competing methods of estimating the noise standard deviation for use in calculating the universal threshold. Donoho and Johnstone (1994) suggest that σ can be estimated using the median absolute deviation of the finest level wavelet coefficients divided by 0.6745. This constant is the median of the absolute values of a standard normal variable and so makes the estimate unbiased under the assumption, $\mathbf{d}^{(J-1)} \sim N(0,1)$.

$$\hat{\sigma}_{\text{mad}} = \frac{\mathtt{median}\left[|\mathbf{d}^{(J-1)} - \mathtt{median}\left(\mathbf{d}^{(J-1)}\right)|\right]}{0.6745}$$

An alternative method for estimating σ is using the sample standard deviation of the wavelets at the finest resolution level.

$$\hat{\sigma}_{\text{sd}} = \sqrt{\frac{1}{\frac{n}{2}-1}\Sigma_{i=1}^{n/2}\left[d_i^{(J-1)} - \bar{d}_i{}^{(J-1)}\right]^2}$$

A small simulation study was carried out to investigate the appropriateness of these two methods.

Comparison of different noise estimation methods

Simulations were carried out with 100 repetitions of 100 values of signal to noise ratio, ranging from 0.1 to 1, corresponding to noise standard deviation 1 to 0.1. The signal to

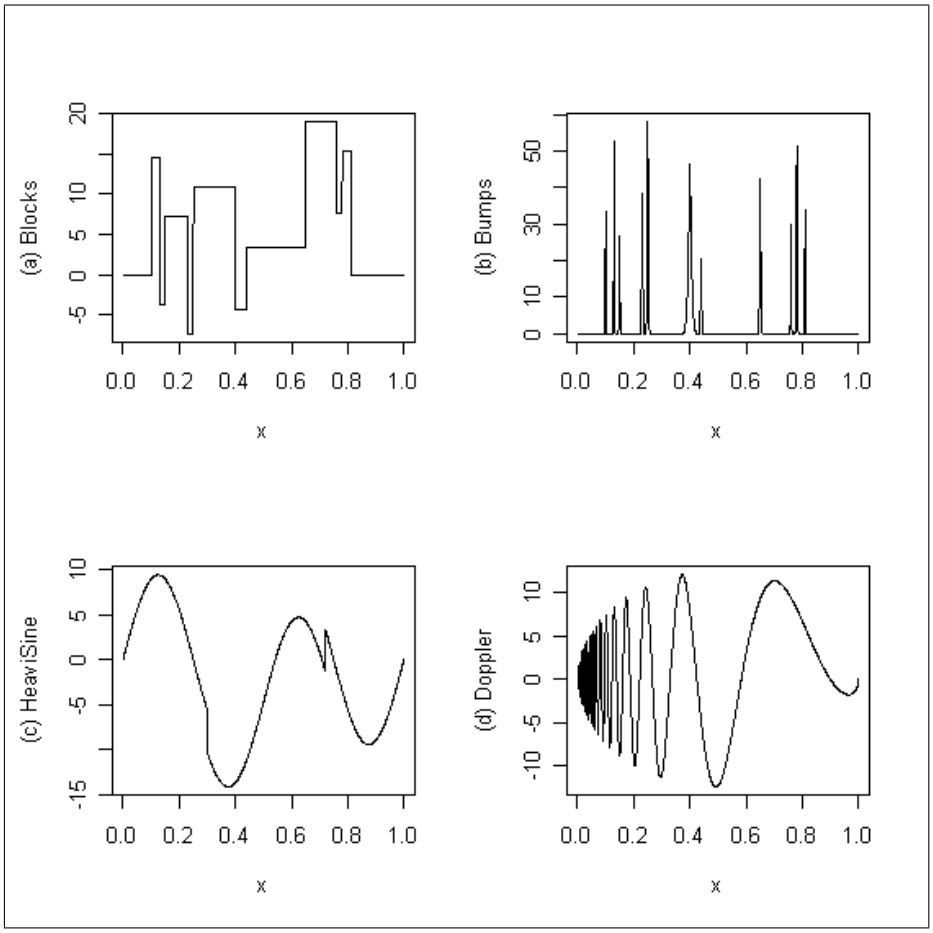


Figure 2.5: The test signals of Donoho and Johnstone (1994).

noise ratio is defined as the reciprocal of the noise standard deviation and in the simulation a signal to noise ratio of 1 corresponds to equal strength of noise and signal. The test signals of Donoho and Johnstone (1994) (see figure 2.5) were used with length $1024 = 2^{10}$.

Various factors which could influence the quality of the noise standard deviation estimate were identified. These were:

- The form of the estimate used. Here the MAD and the standard deviation of the wavelet coefficients are compared.

- The test signal itself: a discontinuous signal with relatively large jumps may cause the noise standard deviation to be overestimated; a high frequency signal may make the noise estimation difficult if it results in coefficients representing signal at the finer resolution level.
- The signal to noise ratio.
- The wavelet transform used — decimated and non-decimated transforms are investigated.
- The wavelet filter used.
- The resolution level of the wavelet coefficients used in the noise estimation.

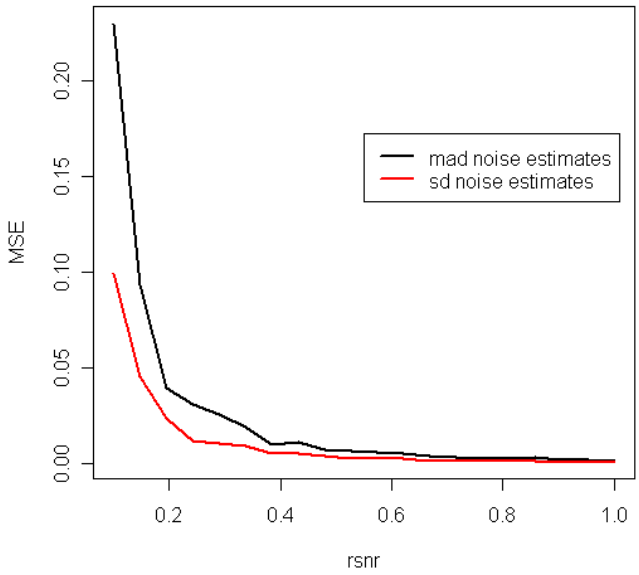
To investigate the effect of filter choice, several different filters were used, but the noise estimation appeared robust to the actual filter. For this reason, only the results using the Daubechies Extremal Phase basis with 10 vanishing moments are reported. The estimates using the finest resolution level coefficients are considered, as noise will be more prominent at the finer resolution levels.

The results for the signal to noise ratio of 0.1 are shown in Table 2.1. This shows that the standard deviation estimate performs better than the MAD estimate for all signals and signal to noise ratios. The biggest differences are observed in the doppler signal when using the decimated transform and the heavi signal for the non-decimated transform.

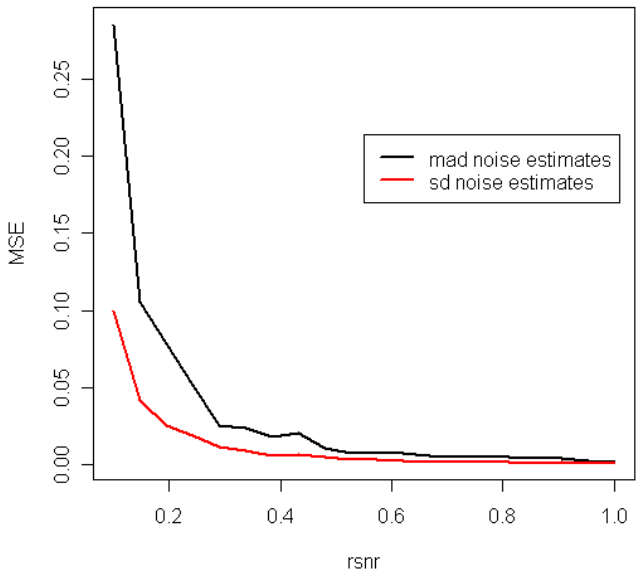
Figure 2.6 shows examples of how the AMSE varies depending upon the SNR for these two signals. It can be seen that the standard deviation method for noise estimation appears to consistently outperform the MAD estimate.

		Test Signal			
		Blocks	Bumps	Heavi	Doppler
DWT	MAD AMSE	0.254	0.254	0.285	0.226
	SD AMSE	0.101	0.101	0.001	0.087
NDWT	MAD AMSE	0.191	0.210	0.191	0.230
	SD AMSE	0.099	0.095	0.093	0.100

Table 2.1: Average mean squared error of the noise estimates using the different methods.



(a) DWT - doppler signal



(b) NDWT - heavi signal

Figure 2.6: AMSE of noise estimates for different signal to noise ratios.

Chapter 3

Modelling Seismic Data

3.1 Introduction

3.1.1 Seismic methodology

Seismic methods are used in the geophysical sciences to explore the composition of underlying rock. These methods utilise the fact that elastic waves travel with different velocities in different rock types, depending on the density of the rock. There are a broad range of seismic applications — ranging from building site investigations and surface environmental studies, to oil and gas exploration, even to detection of water-bearing fracture zones and long-period earthquake seismology.

A reflection seismic setup is considered, as shown in Figure 3.1. The data images in this situation consist of adjacent time series indicating the arrival of sound waves at the geophones, which have been artificially generated from a source, such as a controlled explosion or a sledge hammer. These waves will have been reflected from the interfaces between rock formations with differing physical properties. A key property is the density of the rock through which the sound waves have travelled and this is the property we shall

concentrate on modelling. These traces will provide some insight into the underground structure, with strong reflection events indicating boundaries between rock formations. Some simulated data of this setup will be used in Section 3.3.1 to build a model to predict the underground composition from such traces.

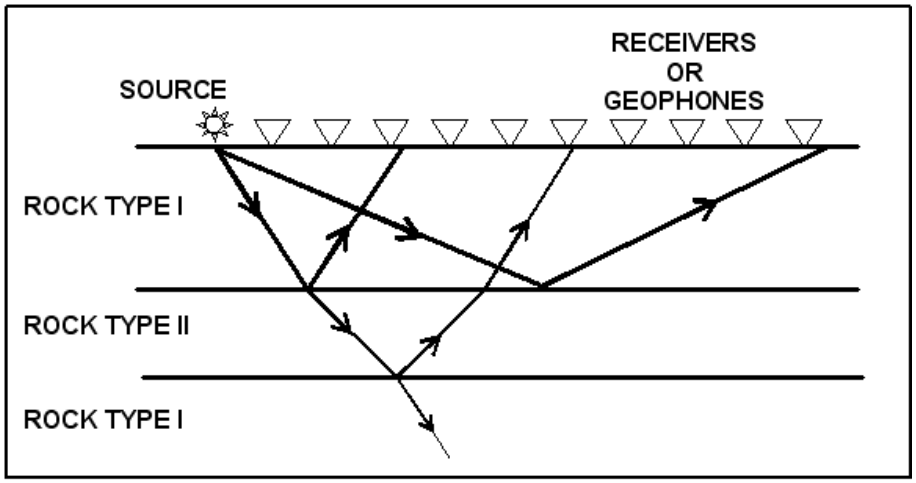


Figure 3.1: Example of reflection seismic setup.

Estimation of the rock density from seismic data often results in an inverse problem, where the solution is unstable to small changes in the data. It is expected that the use of wavelet methods will prove to be useful in avoiding such problems and provide a more reliable rock density reconstruction. See Stein and Wysession (2002) for more details relating to seismic methods.

3.1.2 Properties of the observed signals

There are two main ways in which the composition of underground rock can vary. These can be described as either smooth or sharp changes.

- **Smooth changes** – associated with processes such as compaction, which will produce natural variation in the signal within a rock type. These would represent

low frequency changes in the rock properties.

- **Sharp changes** – these usually occur as the depth of observation is increasing, that is in the vertical direction, and are associated with changes in rock type. The sharp changes in features are what give rise to the reflections and diffractions that are observable in seismic traces. At a sharp change point, the rock properties would change with a higher frequency.

As each of these component changes will be inherent in a seismic trace dataset, the associated seismic traces will have both low and high frequency properties varying over time. It is hope that through the use of wavelet methods, these can be exploited to improve the estimation of the underground lithology.

3.2 Data simulation

3.2.1 Overview

The simulated data was produced from the starting point of a supposed 'known truth' about the geological makeup of an area. This was then convolved with the Ricker transform, which is defined in Section 3.2.2, and a noise process introduced into the observations.

The 'known truth' was a (256×256) dataset, consisting of 256 different observation points along a straight line on the surface. From each of these observation points, the sound trace was recorded over time, representing depth below the point, for 256 time points. A diagram to show the setup is shown in Figure 3.2. We assume there are two clear boundaries dividing rock of differing densities. The pattern is less clear from the simulated data and there is no simple way to map the observations to give an interpretable 'truth'.

An diagram to illustrate the various stages of the data simulation process can be seen in Figure 3.3.

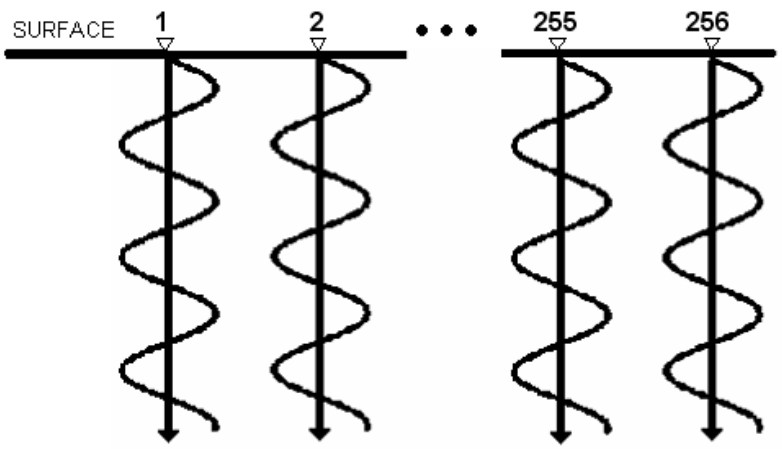


Figure 3.2: Diagram of the simulation setup.

3.2.2 A model for seismic data

A common model used in seismology is that of stacked layers of differing rock types, each with a characteristic density and acoustic velocity. The acoustic impedance, denoted by Z, of the substance is modelled as the product of this characteristic density and acoustic velocity. It is further assumed that different substances carry signals at a characteristic frequency, proportional to the acoustic impedance. The acoustic impedance is given by

$$Z = \rho V + \varepsilon \tag{3.1}$$

where ρ is density, V is the acoustic velocity of the material and ε is an error term.

Where there is an interface of different materials, the reflected and refracted amplitudes can be found using a simple expression. The reflection coefficient, c, relates incident

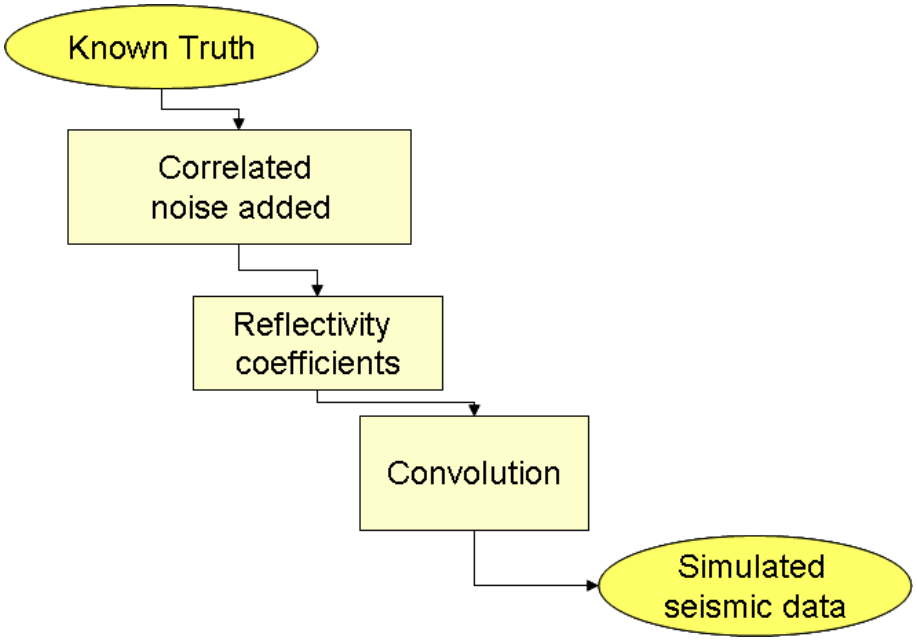


Figure 3.3: Data simulation process.

and reflected trace amplitudes measured in units of pressure. This reflection coefficient is found using the following equation

$$c = \frac{Z_2 - Z_1}{Z_2 + Z_1}$$

where Z_1 and Z_2 are the acoustic impedances for layer 1 and 2 respectively.

The model of recorded seismic data is built by the product of a waveform, which is often taken to be the Ricker or Mexican hat wavelet, with the reflection coefficient series (Robinson and Treitel, 1980). The series of reflection coefficients represents the depth reflections of the material, in the sense that the series will have a spike when the boundary of two substances is reached.

The Ricker wavelet is defined as the second derivative of a Gaussian distribution with zero mean and variance $1/\tau^2$, where τ is related to the dominant frequency of the wavelet. This can be written as

$$\psi(x) = (1 - \tau^2 x^2) \exp^{\frac{-\tau^2 x^2}{2}} .$$

3.2.3 The discrete convolution model

The discrete convolution model (Robinson and Treitel, 1980) can be written as

$$X = \psi \otimes c + \nu$$

where X represents the seismic data, $\otimes$ represents a convolution and ϵ is Gaussian white noise.

Inherent in the calculation of the acoustic impedance, c, is an error structure determined by the nature of the ε term from equation 3.1. As the true values of Z are assumed to be known, this error is usually modelled directly as an additive error term ν. The structure of the noise term ν can be broken down into two components:

$$\nu \quad = \quad \nu_{cor} + \nu_{wn} \tag{3.2}$$

The first component, ν_{cor}, represents correlated noise of a frequency proportional to the acoustic impedance of the material, which is considered in Section 3.2.4. The second component, ν_{wn}, is white noise representing measurement error.

To illustrate the simulation process, figure 3.4 shows the simulation with $\nu_{cov} = 0$ and $\nu_{wn} \sim N(0, \sigma^2)$, with $\sigma = 1$. Figure 3.5 shows the simulation with ν as described in equation 3.2, adding correlated and white noise to the acoustic impedance values. This simulation process was repeated 256 times to obtain a two dimensional simulated dataset.

Comparing Figures 3.4 and 3.5, it can be seen that when correlated noise is added (as generated in Section 3.2.4), the underlying structure of the reflectivity coefficients and hence the synthetic trace become much more erratic.

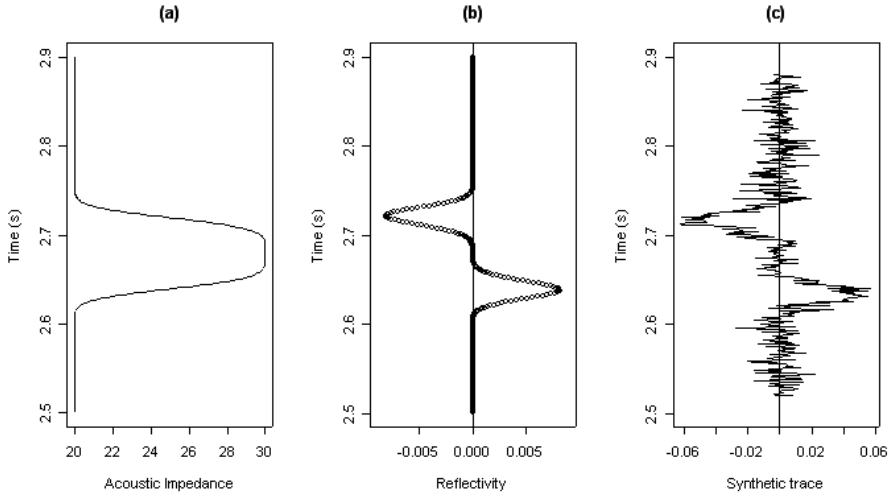


Figure 3.4: (a) Acoustic impedance, (b) Reflection coefficient, (c) Synthetic trace.

3.2.4 Non Gaussian noise processes

Motivation

Noise of a Gaussian nature as considered in section 3.2.3 is a common simplification used when studying seismic processes. It is often assumed that the correlation effects of the signal will be minimal compared to the underlying signal. Zerva and Zhang (1997) have noted that the correlation of the noise process will be affected by the true composition of the underlying rock. For this reason, it is proposed that a noise process with correlation varying with the acoustic impedance of the rock would more accurately simulate a real process. To introduce a varying correlation in the noise process the model, the noise process was simulated from an auto normal process as described in Aykroyd et al. (1996). This process will introduce horizontal as well as vertical correlation into the data simulation process.

A first order neighbourhood will be used in the simulation, which is where the neighbours of the site i are the four orthogonally adjacent sites, as in Table 3.1.

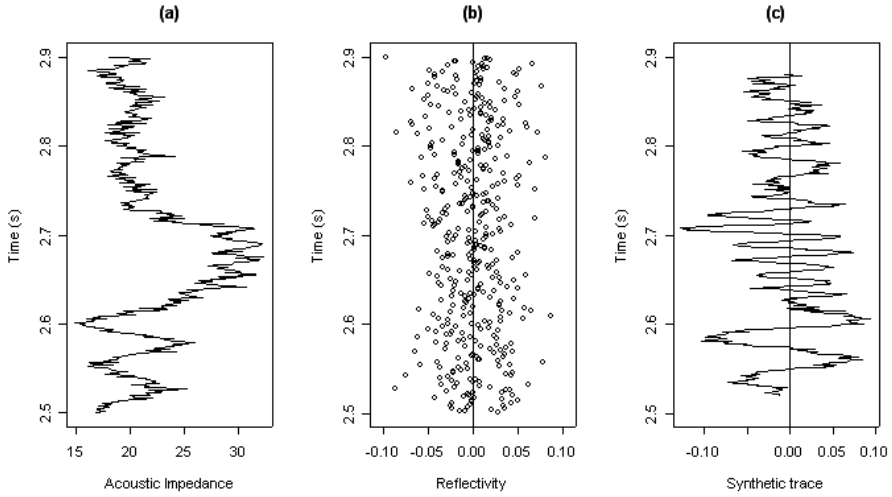


Figure 3.5: (a) Acoustic impedance, (b) Reflection coefficient, (c) Synthetic trace.

	u'	
v	x_i	v'
	u	

Table 3.1: First order neighbourhood.

Let δ be the set of neighbours of point x_i. The auto normal model can now be generated under the assumption that the probability structure at the point x_i depends only on contributions from the neighbouring sites. The noise simulation will sample from an auto Normal distribution, meaning that the parameters of the distribution sampled at x_i will depend on the values at each of its neighbours in δ.

The underlying truth in the following simulations is a sequence of measurements taken at 256 adjacent points along a surface line, each recording 256 time points. All of the traces have a similar structure to that shown in Figure 3.6 (a). The data set is therefore a matrix $\mathbf{Z}$ with 256 rows and 256 columns. This truth was then corrupted by auto normal noise generated via a Gibbs sampler.

Gibbs sampler

Gibbs sampling is an MCMC method which uses an iterative approach to sampling from a probability distribution. The scheme in this case, with periodic boundary conditions and a first order neighbourhood, can be described as:

- Pick a starting matrix $\mathbf{Z}^0_{256\times256} = (\mathbf{z}^0_1, \mathbf{z}^0_2, \ldots, \mathbf{z}^0_{256})$.
- Simulate a new value for each pixel from the conditional distributions, such as

$$\begin{aligned} \mathbf{Z}^1_{1,1} \quad &\text{from } p(\mathbf{Z}_{1,1}|\mathbf{Z}^0_{1,2}, \mathbf{Z}^0_{1,256}, \mathbf{Z}^0_{2,1}, \mathbf{Z}^0_{256,1}) \\ \mathbf{Z}^1_{1,2} \quad &\text{from } p(\mathbf{Z}_{1,2}|\mathbf{Z}^0_{1,3}, \mathbf{Z}^0_{1,1}, \mathbf{Z}^0_{2,2}, \mathbf{Z}^0_{256,2}). \end{aligned}$$

- The matrix is updated using similar conditional distribution for all possible vertical sweeps followed by a full set of horizontal sweeps to produce one full iteration to $\mathbf{Z}^1$. The conditional distribution in each case will be:
  For vertical sweeps,

$$\mathbf{Z}^{t+1}_{x,y} \text{ from } p(\mathbf{Z}_{x,y}|\mathbf{Z}^{t+1}_{x,y-1}, \mathbf{Z}^t_{x,y+1}, \mathbf{Z}^t_{x+1,y}, \mathbf{Z}^t_{x-1,y})$$

  and for horizontal sweeps,

$$\mathbf{Z}^{t+1}_{x,y} \text{ from } p(\mathbf{Z}_{x,y}|\mathbf{Z}^t_{x,y-1}, \mathbf{Z}^t_{x,y+1}, \mathbf{Z}^{t+1}_{x+1,y}, \mathbf{Z}^t_{x-1,y})$$

  where the values of $(x-1), (y-1), (y+1)$ and $(x+1)$ are given in mod 256 to enforce periodic boundary conditions.

Since the distribution is Auto Normal,

$$\mathbf{Z}_{x,y}|\mathbf{Z}_\delta \sim N(a, b\sigma^2)$$

where $a = \bar{\mathbf{Z}}_\delta$, b and σ are constants set equal to 1 in the simulation.

A set of 256 vertical sweeps followed by 256 horizontal sweeps will complete the transition from $\mathbf{Z}^0 \rightarrow \mathbf{Z}^1$.

This is repeated for an initial 'burn-in' period, in this case chosen to be 50 iterations, with the original $i = 1, \ldots, 49$ values for $\mathbf{Z}^i$ being discarded. The convergence of the simulation was monitored over the iterations to ensure a sensible estimate for the acoustic impedance truth was obtained. Further details on convergence can be found in Green and Han (1992).

3.2.5 Diagnostics of simulated data

Figure 3.6(a) shows a sample acoustic impedance series with auto normal noise added as described in Section 3.2.4 above. Figure 3.6(b) is the wavelet decomposition of this series and shows that there is different activity in the central region of the Depth detectable at resolution levels 4,5 and 6.

This data will then be input into the deconvolution model to create a set of example recorded traces, which can then be used in the model building process.

3.3 Modelling method

3.3.1 Modelling framework

The data contains a noise process with varying frequency properties and so a wavelet approach will be used to model the known truth. This will then be applied to data with a different structure to test the models predictive power and to assess if the model is over fitted.

The underlying signal is a time series of 256 observations of a synthetic trace. It is also

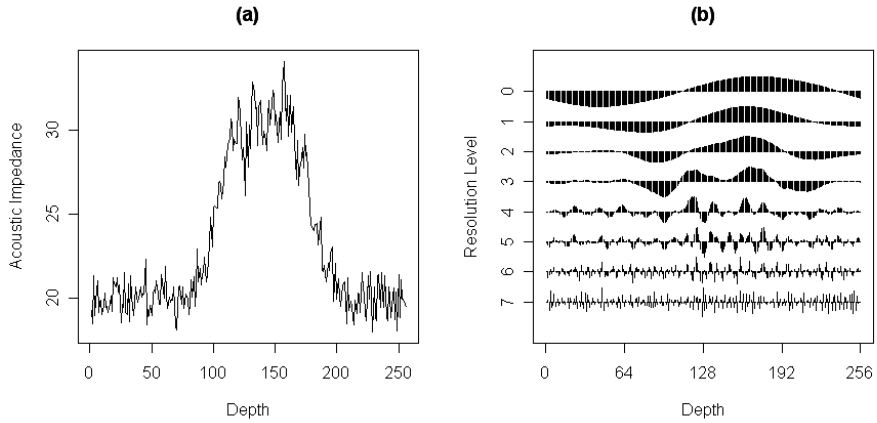


Figure 3.6: (a) Example of simulated acoustic impedance trace and (b) its wavelet decomposition using the NDWT DaubEx(2).

assumed that the underlying truth is known for one of the traces, practically this could be obtained by digging a bore hole down, after observing the trace. The initial modelling framework is:

- take the wavelet decomposition of the synthetic trace using a wavelet basis with short support, here the Daubechies Extremal Phase basis with 2 vanishing moments is used.
- pre-process the wavelet coefficients using some suitable function, the *transfer function*, which will be explained further in Section 3.4.
- build a linear regression model using the known acoustic impedance truth as the dependent variable, with the processed wavelet coefficients as explanatory variables.

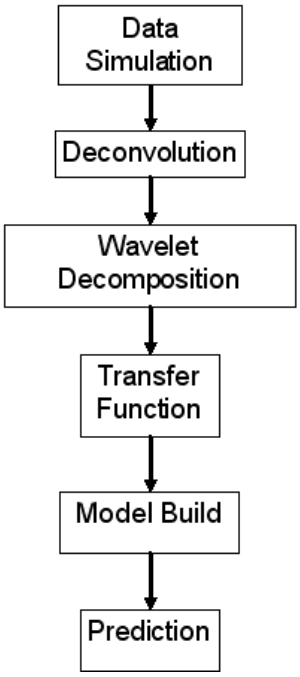


Figure 3.7: Overview of the modelling process.

3.3.2 Prediction

The model can now be used to predict the values of the true acoustic impedance from the realisations of the synthetic trace. For each predicted point, a $100(1-\alpha)\%$ confidence interval for the fitted value will be given by

$$\hat{Z} - t_{(\frac{\alpha}{2},n-2)} \times se_{fit} < Z < \hat{Z} + t_{(\frac{\alpha}{2},n-2)} \times se_{fit}$$

where $t_{(\frac{\alpha}{2},n-2)}$ is the $\frac{\alpha}{2}$th quantile of the t distribution with $(n-2)$ degrees of freedom, n is the number of observations and se_{fit} is the standard error of the fit. This will produce a point wise confidence interval.

As shown in Figure 3.4, the synthetic trace is the final result of a multi-stage simulation process, which starts with the acoustic impedance, then adds a correlated noise process,

the reflectivity coefficients are then found and finally the trace is obtained after convolution with a Ricker function.

3.3.3 Deconvolution

The convolution in the data simulation process in effect takes a snapshot of the data at one particular frequency, that of the Ricker function used in the convolution step. This is in effect destroying the multi-frequency nature of the data that it was hoped the model would exploit. To reconstruct some of this information, an attempt to deconvolve the data was made.

The method for the deconvolution was to undo the convolution using a best guess at the actual filter frequency. The actual frequency used in the data simulation was 50 Hz. A range of filter frequencies was investigated and the effects on the reflectivity coefficients and then the resulting predictions are shown in Figure 3.8. The graphs show that the error in the prediction from the model is flat across all the frequencies investigated. This means that the model is relatively robust to the gaps in the knowledge about the convolution frequency and an estimate of this frequency will prove adequate. It can also be seen from the graph that the errors in the reflectivity coefficients is at a comparable level if your guess at the actual frequency is within ± 20Hz of the true frequency.

In practice, estimating the frequency of the deconvolution filter should not pose a problem since this should be known from the initial experimental setup and the apparatus used. However, for use in data processing without the knowledge of the exact frequency originally used, the modelling methods are robust to the choice of deconvolution filter.

Accuracy of reconstructing the reflectivity coefficients

The black circular points in Figure 3.8 show the average mean squared error (AMSE) of the reconstructions over 256 realisations. The frequency of the Ricker wavelet used in the

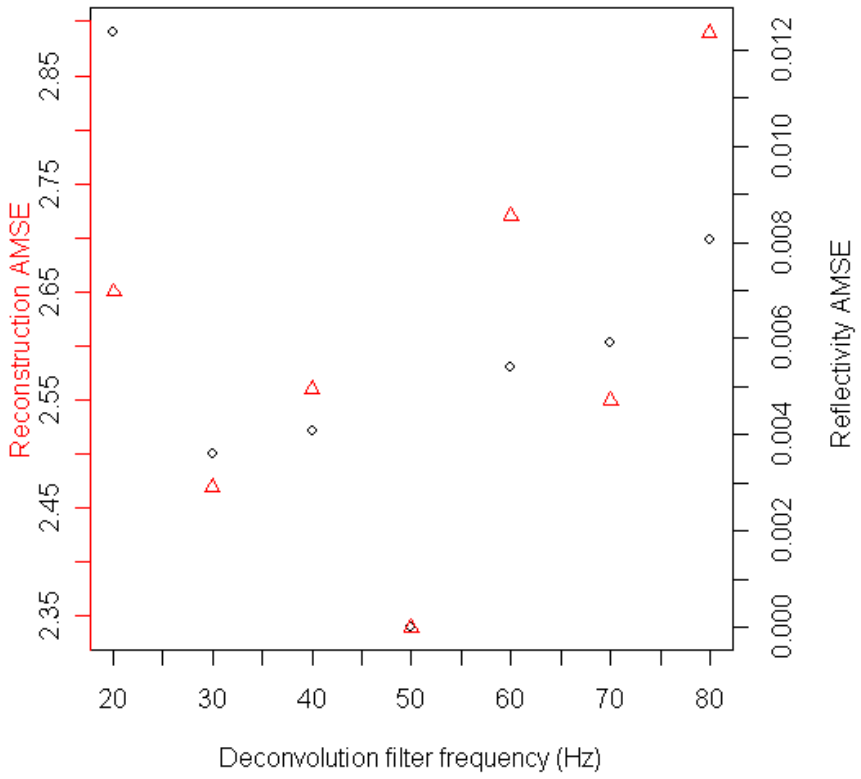


Figure 3.8: Comparison of the errors for unknown filter frequency.

simulation process was 50 Hz, and so the range of investigation was extended to 20 to 80 Hz. The AMSE can be seen to be minimised when the deconvolution filter frequency was also 50Hz, where a perfect reconstruction was possible.

To investigate how the final reconstructions are affected by error in the deconcolution frequency, the running median transfer function was used to build a wavelet model, this choice will be explained in Section 3.4. The triangular points show the AMSE of the acoustic impedance reconstructions using the regression model. It can be seen that the deconvolution filter frequency has little effect on the predictive success of the model when the value used is within ± 20Hz of the true value. Outside this range, as the goal of the

model is to minimise the reconstruction error, it would appear that ‘guessing’ a lower value for the filter frequency would give better results.

Effect of deconvolution on prediction

After deconvolving the data, the error in the prediction is decreased with the mean squared error of reconstruction reducing from 4.32 to 2.34.

It appears as though undergoing some sort of deconvolution process, regardless of the frequency of the filter used, reinstated some of the localised behaviour and frequency shifts that it was hoped the wavelet modelling approach would be able to exploit. The results from the previous section suggest that knowing the precise frequency for deconvolution is not essential, therefore it is possible to deconvolve initially in practise when the true frequency is not known.

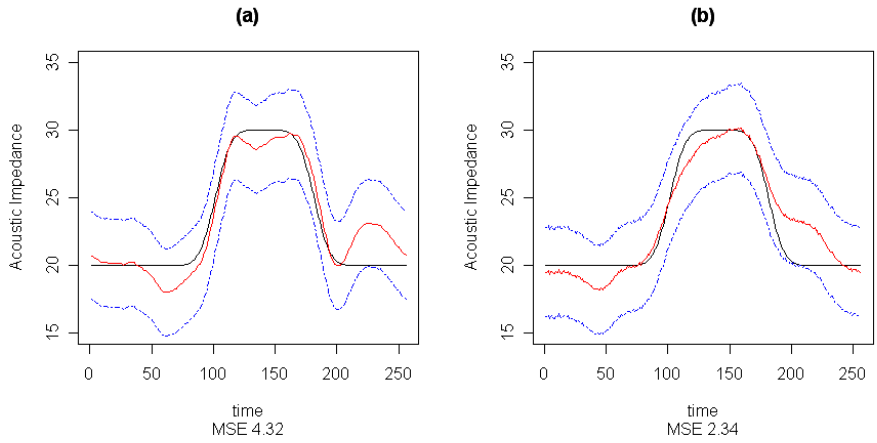


Figure 3.9: Reconstructions using: (a) the raw data; (b) a deconvolved version of the data.

3.4 Transfer function choice

The wavelet coefficients require some pre-processing in order to exploit the desired properties of the data. This is achieved by applying a transfer function to the wavelet coefficients before the model building stage. The transfer function will:

- be specific to the particular setting and nature of the data.
- be related to the structure observed in the wavelet decomposition.
- encode the features of interest.
- identify these characteristics with the response variable.

There will be many possible choices for the transfer function and it is possible that alternatives will produce vastly different results.

3.4.1 Suggested transfer functions

There are several possible transfer functions which could be used. Some of these are:

- Identity function, that is using the raw coefficients, $\{d_{jk}\}$.
- Moving average of the wavelet coefficients of length a, that is

$$\frac{1}{a}\sum_{l=1}^{a} d_{j,k^*} \quad \text{where} \quad k^* = (k+1-l) \bmod 256.$$

- Running median with window l, that is

$$\text{median}\,(d_{j,k}, \ldots, d_{j,k*}) \quad \text{with } k^* \text{ as defined above.}$$

- Absolute value , $\{|d_{jk}|\}$.

There are other signal filters in the literature that could be used in this time-domain setting. An example of such a filter is the Butterworth filter (Butterworth, 1930) which is designed to have a frequency response which is as flat as possible. Another example is the Chebyshev filter which minimises the error between the idealised filter characteristics and the actual over the range of the filter. Further details can be found in Williams and Taylor (1995) and Oppenheim and Schafer (1999).

3.4.2 Diagnostics of data to motivate transfer function

Figure 3.10 shows the observable patterns in the wavelet coefficients for each of the different transfer functions suggested in Section 3.4.1. From these initial plots it can be seen that, at resolution levels 3,4 and 5, all of the resulting coefficients pick up the differences in the trace in the central region of the Time dimension, where the acoustic impedance is higher. At the finer resolution levels, the difference is less noticeable. Looking at Figure 3.10 (c), a reduction in the magnitude of the coefficients in this central region is more observable than with the other suggested transfer functions.

3.5 Application of the proposed method

The modelling method described in section 3.3.1 was applied to the simulated data set. The reconstruction was compared after using several choices of transfer function and the results are shown in figure 3.11.

3.5.1 Sample reconstructions

The running median transfer function was chosen for the seismic model due to the better performance of the resulting model in terms of the mean squared error of the

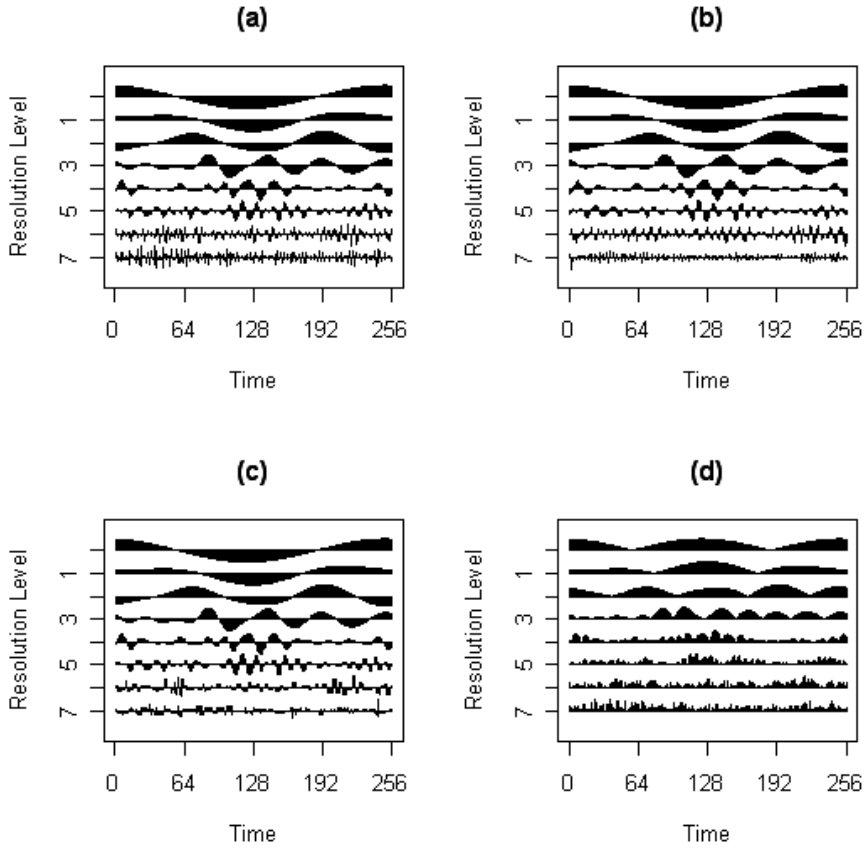


Figure 3.10: Plots of the transformed wavelet coefficients: (a) Identity, (b) Moving average with window 5, (c) Running median with window 5 and (d)Absolute value.

reconstructions. The resulting reconstructions of other transfer functions can be seen below in Table 3.2.

3.5.2 Application to different simulated data

All of the datasets used so far have been a simple set up with the acoustic impedance starting low for 80 time points, getting high for a period of 125 time points, before returning to the initial low value. It was originally assumed that the cross section along

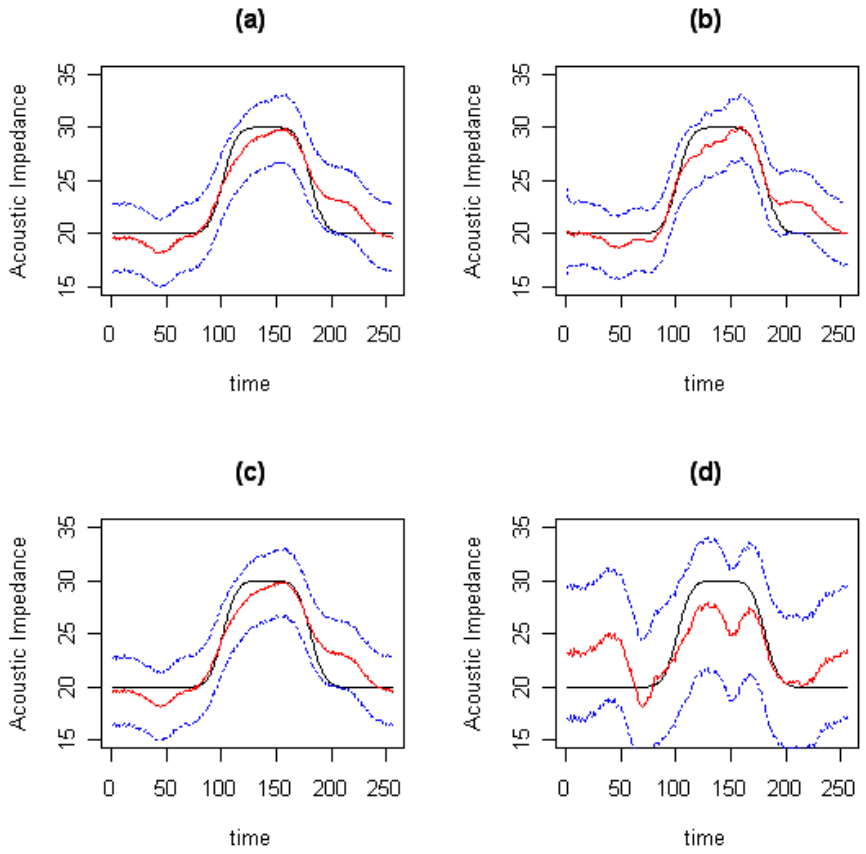


Figure 3.11: Reconstruction using different transfer functions: (a) Identity, (b)Moving average, (c)Running median and (d)Absolute value.

the line of interest would have a uniform configuration of rock types.

The next stage in testing the model performance was to demonstrate that it did not require this exact structure in order to be able to provide reasonable predictions. The structure was changed to have non-linear boundaries between rock types along the investigated region.

Figure 3.12 shows the noisy truth of the new dataset, with some examples of the acoustic impedance traces along the surface. This different structure means that the periodic

Transfer function	MSE
Identity	2.61
Moving average	2.90
Running median	2.34
Absolute value	11.60

Table 3.2: Mean squared error of the reconstructions.

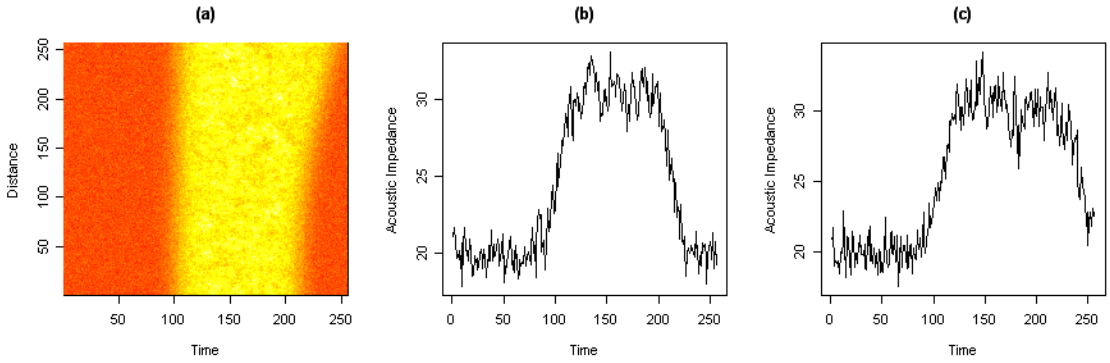


Figure 3.12: (a)image of an alternative 256×256 simulated truth, with non-linear rock type boundaries (b) and (c) are two acoustic impedance examples from this simulated set.

boundary conditions used in the wavelet decomposition may be more questionable. Figure 3.14 shows that the modelling method still appears to work, with the reconstruction MSE of 3.08 being slightly higher than for the previous dataset.

3.6 Incorporating neighbouring traces

In geological applications, it is often noted that adjacent traces have similar acoustic impedance patterns. It was considered that the traces obtained from the two neighbouring sites may improve the predictions obtained from the model. The idea behind

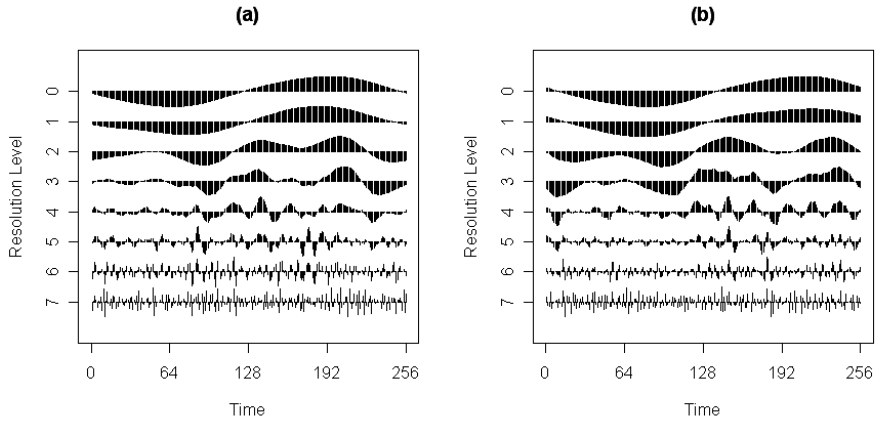


Figure 3.13: Stationary wavelet transform of the sample truths from Figure 3.12 (b) and (c) respectively using Daubechies Extremal Phase basis with 2 vanishing moments.

introducing neighbouring traces into the model was to investigate whether information for surrounding points would improve the predictions. This was particularly of interest when there was an underlying trend in the acoustic impedance boundaries, as is the case in the alternative dataset which has a gradual increase in the depth of the acoustic impedance boundaries. From a practical point, geophysical exploration is usually carried out on a grid and incorporating neighbouring traces would further increase the model's flexibility.

Consider the set of deconvolved noisy simulated reflectivity coefficients, $\mathbf{Z} = (Z_1, \ldots, Z_{256})$, with each of the recorded traces being of length 256, generated to represent an underlying true acoustic impedance structure, $\mathbf{X} = (X_1, \ldots, X_{256})$. Let the wavelet coefficients of each of the Z_i be denoted by $\{d^i_{jk}\}$, then for each value of $i = 1, \ldots, 256$, there is a corresponding set of wavelet coefficients $\{d^i_{jk}\}$, with $j = 1, \ldots, 8$ and $k = 1, \ldots, 256$.

The chosen transfer function, the running median with window length 5, is then applied to each of the $\{d^i_{jk}\}$ and these processed coefficients will be denoted by $\{\tilde{d}^i_{jk}\}$.

The underlying truth for the one of the locations along the line of interest, X_p, is assumed

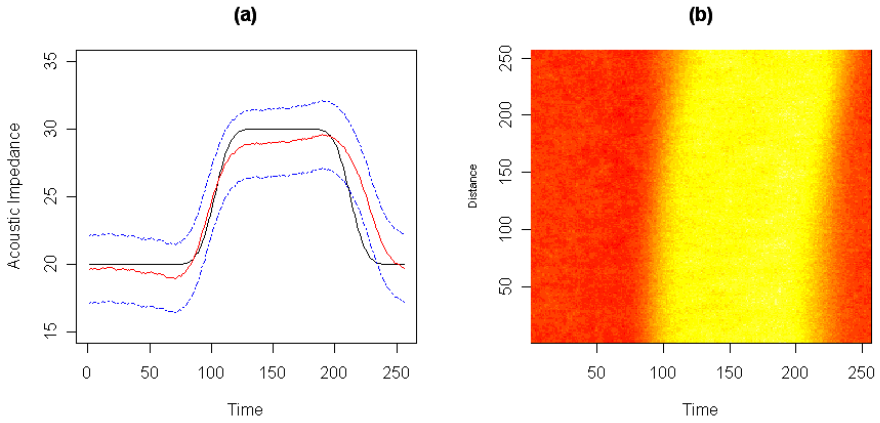


Figure 3.14: (a)Representative reconstruction of the alternative dataset (b)Mean reconstruction of the whole image.

to be known and a linear regression model can then be built using the processed wavelet coefficient series as explanatory variables. This produces a set of k observations of each of the $\tilde{d}_j^i$'s, which are then used to build a model of the form:

$$X_p = \sum_j \alpha_j^{(p-1)} \tilde{d}_j^{(p-1)} + \alpha_j^p \tilde{d}_j^p + \alpha_j^{(p+1)} \tilde{d}_j^{(p+1)}$$

where $j = 1, \ldots, 8$ corresponding to the 8 resolution levels. This model is then used to predict each of the $X_2, \ldots, X_{255}$ from the corresponding series of processed wavelet coefficients. Note that predictions are only made for traces with neighbouring traces, that is no predictions are produced for the X_1 and $X_2 56$ traces.

3.6.1 Original Dataset

Figure 3.15 shows the reconstruction of the acoustic impedance, using the recorded traces, based on this neighbour modelling method. Comparing this to Figure 3.9(a), it can be seen that a similar 'bump' can be observed after the true acoustic impedance returns to its lower

value. However, comparing the average mean squared error of the reconstructions using both methods, the neighbouring traces method performs better, with an AMSE of 3.49 compared to that of 2.34 for the sole trace method.

Figure 3.16 shows the reconstruction of the Acoustic impedance, using the deconvolved data estimates for the reflectivity coefficients, based on this neighbour modelling method. This reconstruction appears to be similar to that using only one trace (Figure 3.9(b)) but again the neighbouring traces method reduces the AMSE of the reconstruction from 2.34 to 2.11.

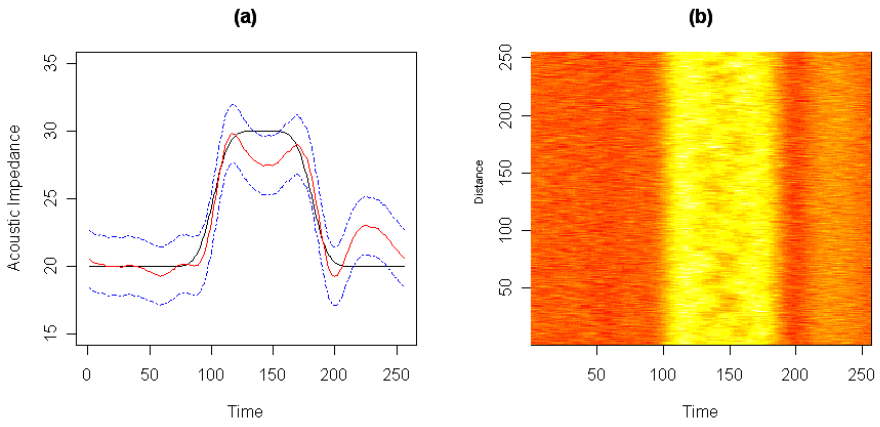


Figure 3.15: Neighbouring Traces reconstruction using recorded traces, (a) representative reconstruction, (b) image reconstruction.

3.6.2 Alternative Dataset

Figure 3.17 shows the reconstruction of the Acoustic impedance, using the recorded traces, based on this neighbour modelling method. Figure 3.18 shows the reconstruction of the Acoustic impedance, using the deconvolved data estimates for the reflectivity coefficients, based on this neighbour modelling method. The reconstructions using both the data and the deconvolved reflectivity coefficients appear closer to the underlying truth

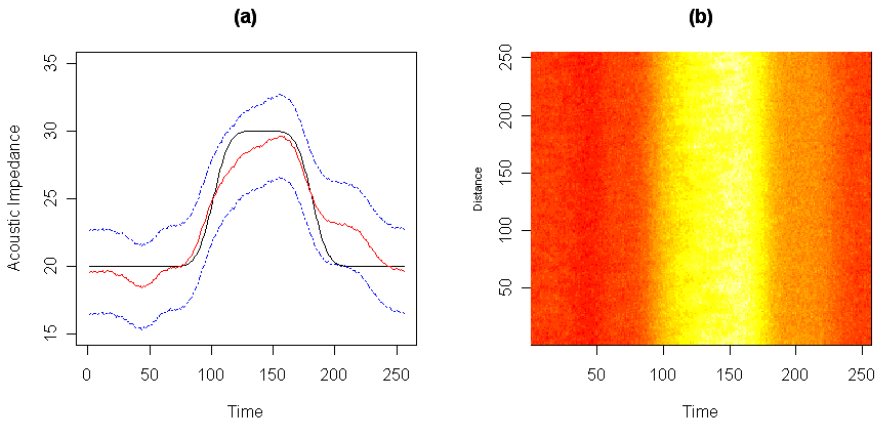


Figure 3.16: Neighbouring Traces reconstruction using deconvolved reflectivity coefficients, (a) representative reconstruction, (b) image reconstruction.

for this dataset. However, the model using the deconvolved version of the data performs slightly better in terms of the AMSE of the reconstruction, but with less percentage improvement than in the original case. The overall AMSE is also reduced, suggesting that when there is a changing structure in the rock decomposition, using the neighbours in the modelling improves the overall fit.

Model data	Original Dataset	Alternative Dataset
Reflectivity coefficients	2.11	1.16
Recorded trace	3.49	1.80
No neighbours reflectivity	2.34	3.08

Table 3.3: Average mean squared error of the reconstructions using the nearest neighbours.

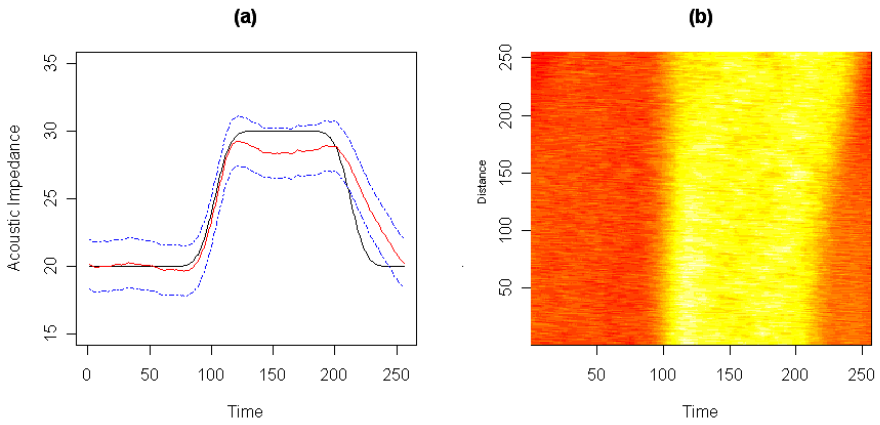


Figure 3.17: Neighbouring Traces reconstruction using recorded traces - alternative dataset, (a) representative reconstruction, (b) image reconstruction.

3.7 Conclusions and further work

Modelling using transformed wavelet coefficients can provide an insight into the underlying frequency structure when used to model acoustic impedance in differing rock types. The drawback to this modelling method is that it requires some instance of the true acoustic impedance to be known in order to train the model.

Convolution, which occurs in a standard reflection seismic setup, can mask some of the frequency structure contained in the data. However, deconvolution seems to restore some of these properties and is robust to the actual frequency of the deconvolution filter.

Using neighbouring traces in order to train the wavelet models on more explanatory variables increases the accuracy of the reconstruction. This improvement is reproduced during replication of the data simulation, suggesting that this is not due to over fitting.

Some of the reconstructions of the original dataset have demonstrated a ‘bump’ downstream of the feature. It may be of interest to analyse the data ‘backward’ to see whether the bump appears in both cases. In this case improvements maybe observed if

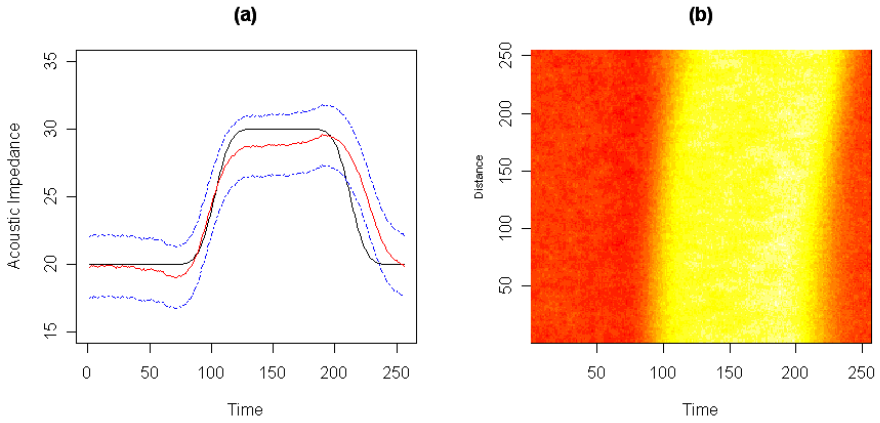


Figure 3.18: Neighbouring Traces reconstruction using deconvolved reflectivity coefficients - alternative dataset, (a) representative reconstruction, (b) image reconstruction.

analysis in both direction was undertaken and the average used for prediction.

Various methods exist in the literature for the analysis of seismic data, from spectral methods (Chakraborty and Okaya, 1995) to fuzzy clustering (Peijie et al., 2006). However, the applications vary and there is no direct comparison of the performance of the techniques developed to these existing methodologies.

Chapter 4

Tomographic Data Modelling

4.1 Introduction to tomography

Electrical tomography techniques provide a cheap and non-invasive approach to the study of static and dynamic processes. Such techniques are widely used in geophysical, industrial and medical investigations. The key common feature of all tomographic techniques is that measurements are taken outside or on the boundary of a region with the aim of describing what is happening within the region. In industrial applications of electrical tomography, multiple voltages are recorded between electrodes attached to the boundary of, for example, a pipe. The most usual first step of the analysis is then to reconstruct the conductivity distribution within the pipe. The most commonly used approaches to reconstruction are based on domain discretization, for example using the finite-element method, leading to ill-posed inverse problems. Usually, such problems are ill posed because there are multiple solutions and the solutions do not depend continuously on the observed data. Stable solution then requires regularization. Even if reliable reconstruction is possible it only provides an image representing the conditions within the pipe. Although this is useful for process visualisation, for the automatic control such an image is at best unnecessary, and will require further post-processing to allow

control parameters to be obtained.

There is a growing sentiment (Stitt and James, 2003) that in a field application the ambitions may be far more limited than in the research environment. That is simple questions such as "is there or is there not a problem?" are more important than obtaining high quality flow and phase patterns within the vessel. Also, that the aim should be to reveal process behaviour in process terms rather than industrial process tomography terms (Hoyle, 2004). One interpretation of these suggestions is that control parameter estimation, rather than process visualization, is the more appropriate output of a data analysis in many real situations. In particular, there is a need for sensing systems, modelling and algorithms that are simple, fast and can operate largely unsupervised.

4.1.1 Flow pattern terminology

When two different components are flowing through a pipe, the simplest model of the flow assumes that the minor component is distributed uniformly within the main component. However, this will not always be the case and different flow patterns will produce different expected tomography measurements. Thus to identify any desired anomalies in the flow, the flow pattern must be considered.

Non-uniform flow patterns may occur in flows consisting of any combination of liquid, solid or gas. For illustrative purposes, the flow of a mixture of gas and liquid is considered. Such situations have a high density difference between the two components and are readily observed in industry.

To describe the flow pattern, it must be acknowledged that there are many possibilities due to the surface tension effects. In practice, the flow type is often classified into one of various 'flow regimes' which provide a qualitative and subjective description of the flow pattern.

The typical flow regimes, shown in Figure 4.1, for a vertical pipe are:

- bubble flow. This is where the gas bubbles are dispersed throughout the liquid.
- slug flow. This occurs when the bubbles coalesce and become large so that they are almost as wide as the pipe.
- churn flow. As the velocity of slug flow bubbles increases, they start to break down. This creates an unstable regime with both wide and small bubbles.
- annular flow. The liquid flows on the walls of the pipe and the gas flows,with small liquid droplets, in the centre.

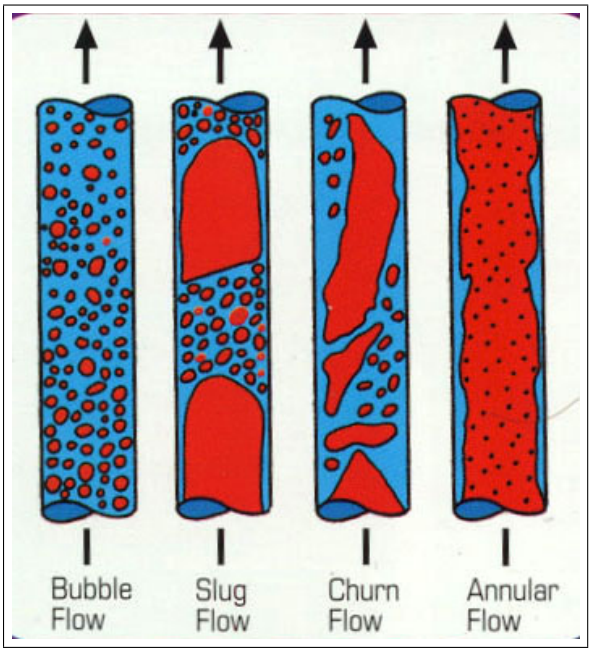


Figure 4.1: A visual representation of some typical flow regimes.

4.1.2 Experimental setup

Consider the flow of a gas through a liquid in a section of vertical pipe. The gas enters at the bottom of the section of pipe under pressure and travels rapidly up the length of the

pipe. The gas fraction and bubble size are controlled by the inlet size and by the input pressure. To control process efficiency it is important to monitor the flow regime, and to adjust the input parameters accordingly. In the following simulation study, bubble, churn and an intermediate flow between these will be considered. It is assumed that the gas has conductivity 1 Ωm and the liquid conductivity 2 Ωm. The key part of the simulation is to generate spatial patterns for the bubbles that evolve temporally. These will in turn define the conductivity distributions.

In electrical tomography, for given conductivity distributions the boundary voltages are found using Maxwell's equations, and appropriate boundary conditions. This is the forward or direct problem. The forward problem is solved numerically using the finite-element method (FEM). For examples of FEM-based approaches see West et al. (2003, 2004).

The data simulation scheme used later in this chapter is motivated by the widely used 'reference protocol' for an eight-electrode electrical tomography system. This is when the electrodes are equally spaced in a ring around a cylindrical pipe. In this protocol a fixed current is passed between a common reference electrode and each of the other electrodes in turn, hence producing seven current patterns. For each current pattern an induced potential field is created within the pipe which depends on the current pattern, and upon the conductivity distribution within the pipe. This potential field is probed by taking multiple voltage readings between the reference electrode and each of the other electrodes. For each current pattern seven voltages are recorded leading to 49 measurements at each time point. Further, the process is allowed to evolve for 256 time points. Once noise-free voltages are obtained uncorrelated Gaussian noise is added to yield the simulated dataset. Hence the full dataset forms 7-by-7 electrode-pair time series of length 256.

4.1.3 Local frequency properties of recorded data

The signals from each of the flow regimes exhibited extremely different behaviour, which can be explained in terms of the frequency of the change in the measurements. The bubble flow with its small pockets of gas in liquid produced more rapidly varying measurements than the churn flow, with the intermediate flow lying somewhere in between (see Figure 4.2). It was anticipated that this difference in the traces would be detectable at the various resolution levels of the wavelet decomposition (Figure 4.3), which contains information about the detail of the signals at differing scales.

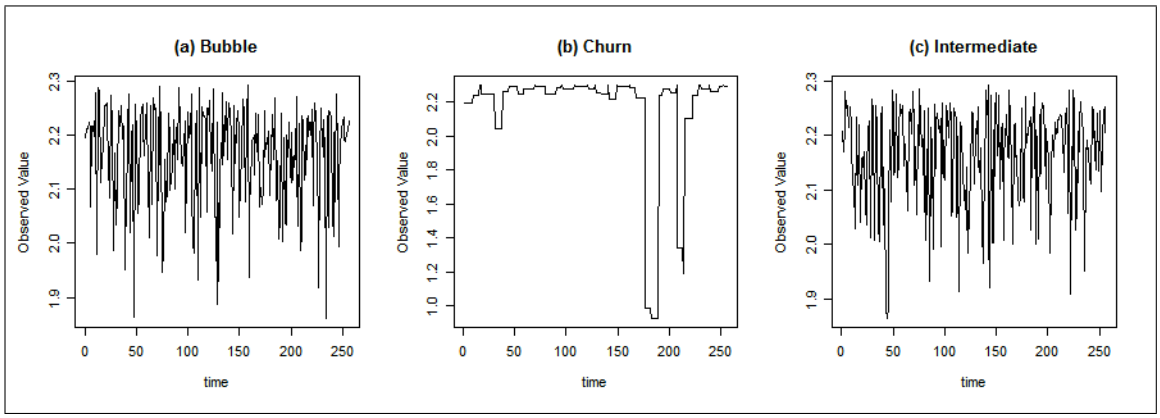


Figure 4.2: Sample simulated series from each of the flow regimes.

Looking at the plots of the non-decimated transforms (Figure 4.3) of the measurement data, it can be seen that at the finest resolution level, there is a lot of activity for the bubble flow and considerably less for the churn flow. Any fine scale activity in the churn flow seems to correspond to the appearance of large bubbles. Looking at the coarse scale activity, it can be seen that most of the activity of the churn flow occurs here, with much less activity from the bubble flow.

This indicates a difference in the high frequency activity of the signals which may be easily detectable using the wavelet transform.

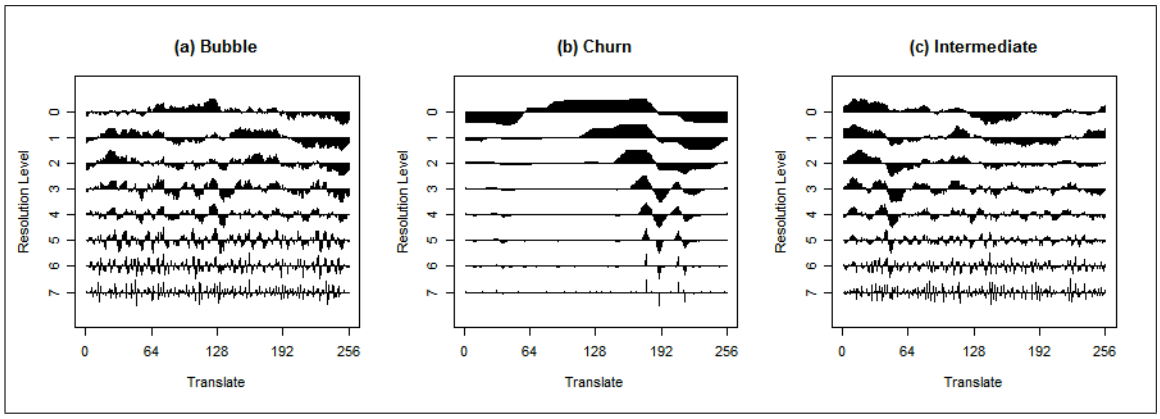


Figure 4.3: Non-decimated wavelet decomposition of the corresponding time series from Figure 4.2.

4.2 Data simulation

4.2.1 Method

The simulation of the recorded voltages is possible as, if the resistivity of the pipe contents is assumed, then boundary voltages can be calculated through the solution of Maxwell's equations. In practice this is done numerically, using the finite element method. The problem is well posed and voltages can be obtained at least to the accuracy of measurements.

In the direct problem, the resistivity distribution ρ is specified. It is sometimes more convenient to make use of the conductivity $\sigma = 1/\rho$. An assumption in the simulation scheme is that resistivity is constant across pixels. The electric-field potential is required at points of the domain boundary from which potential differences (voltages) $\mathcal{U} \in \mathbb{R}^M$ can be calculated.

This is a well-studied and supported mathematical problem relating the unknown conductivity distribution and the observed voltage measurements. If the conductivity

distribution is given then voltages on the boundary can be found using Maxwell's equations, and appropriate boundary conditions. In practice this is usually done numerically, using the finite element method (FEM) (see Vauhkonen et al. (2001) for details).

4.2.2 Some mathematical background

Within a domain, Maxwell's equations can be condensed to the form

$$\nabla(\sigma\nabla\phi) = 0$$

for the conductivity vector σ.

The boundary of the domain will comprise of electrodes $E_k(k = 1, \ldots, 8)$, leading to the following boundary conditions (Cheng et al., 1989).

$$\left(\phi + \zeta_k\sigma\frac{\partial\phi}{\partial n}\right)\Big|_{E_k} = U_k \qquad k = 1, 2, \ldots, K$$
$$\int_{E_k} \sigma\frac{\partial\phi}{\partial n}dS = I_k \qquad k = 1, 2, \ldots, K$$
$$\sigma\frac{\partial\phi}{\partial n}\Big|_{\partial\Omega\setminus\bigcup_{k=1}^{K} E_k} = 0$$

Within the area being considered Maxwell's equations can be used to produce the following elliptic partial differential equation relating the spatial conductivity distribution σ to the electrical potential, ϕ (give a reference)

$$\nabla\cdot(\sigma\nabla\phi) = 0. \tag{4.1}$$

On the boundary, where the electrodes are located, there are additional boundary conditions:

$$\left(\phi + \zeta\sigma\,\frac{\partial\phi}{\partial\underline{n}}\right)\Big|_{E_k} = U_k, \qquad \int_{E_k}\sigma\frac{\partial\phi}{\partial\underline{n}}dS = I_k, \quad k = 1, 2, \cdots K, \tag{4.2}$$

on the electrodes, E_k $(k = 1, 2, \ldots, K)$, and on the insulating boundaries between electrodes

$$\sigma \frac{\partial \phi}{\partial \underline{n}} = 0, \tag{4.3}$$

where ζ is the contact impedances of the electrodes, U_k denotes the potential on the k^{th} electrode E_k, I_k the current at electrode E_k and $\underline{n}$ is the outward unit normal of the boundary. The boundary conditions simply say that the potential measured at an electrode is higher than on the inside surface of the electrode due to the contact impedance, that the total current integrated over the surface of the electrode is equal to the specified current, and that no current passes across the insulating boundaries.

Here current passes between a reference electrode and each of the other K electrodes in turn, hence producing $K - 1$ current patterns and leading to $m = (K - 1)^2$ measured voltages. These voltage measurements are obtained by solving this system of equations. It is not possible to obtain the voltages explicitly so instead the finite element method (FEM) can be used and here calculation have been done in Matlab using the software package EIDORS-2D (Vauhkonen et al. (2001)) based on code developed by West et al. (2004).

To produce the data used here a conductivity distribution is generated by selecting bubble locations randomly across the region. The number of bubbles and their sizes depend on the regime, that is bubble or churn, within the study area. Bubble flow has many small bubbles which pass the electrodes quickly. In contrast churn has fewer but larger bubbles which occupy a larger cross-sectional area and take longer to pass the electrode position. These conductivity distributions are passed to the Matlab code which then gives calculated voltages, that is ones without measurement error. The data measurements are then obtained from the calculated voltages by adding independent gaussian noise.

4.2.3 Control parameters

There are several parameters that may be altered to simulate different flow regimes. These will include:

- Gas volume — the proportion of gas to liquid was kept constant over time.
- Conductivity of the component fluids — this was set to be constant with a mixture of two fluids with conductivity of $1\,\Omega m$ and $2\,\Omega m$ respectively.
- Bubble size — determines the nature of the flow. Many small bubbles or a few larger bubbles would be examples of different types of flow.
- Length of bubble. This would reflect the likelihood of the bubble remaining in the same place as time evolves.

4.3 Modelling using frequency characteristics

4.3.1 Introduction

The idea is to use the coefficients $\{d_{jk}\}$ of the non-decimated wavelet transform in the model building process. The reason for this is that different frequency properties will manifest themselves as coefficients of increased magnitude at different resolution levels. For example, high frequency characteristics will mean a lot of activity at the finer resolution levels, while low frequency characteristics will be represented at coarser levels. One of the aims of the modelling process is to detect changes in flow regimes and for this reason, Haar wavelets were used. Longer filters will incorporate information over longer time windows in the coefficients and may affect the model's ability to detect the flow regime change points.

Initially for the tomography data, a combination of bubble and churn flows are considered (Section 4.3.3). The nature of the data suggests a logistic regression model will be appropriate. A more challenging problem is given by data having some characteristics of both churn and bubble flow (called the intermediate flow) being considered. For these data, a multinomial logistic regression model was appropriate (Section 4.3.5).

In both modelling approaches, working with the raw coefficients gave poor results in terms of flow regime prediction and so a *transfer function* (Section 3.4) was incorporated into the modelling process, which is illustrated in Figure 4.4.

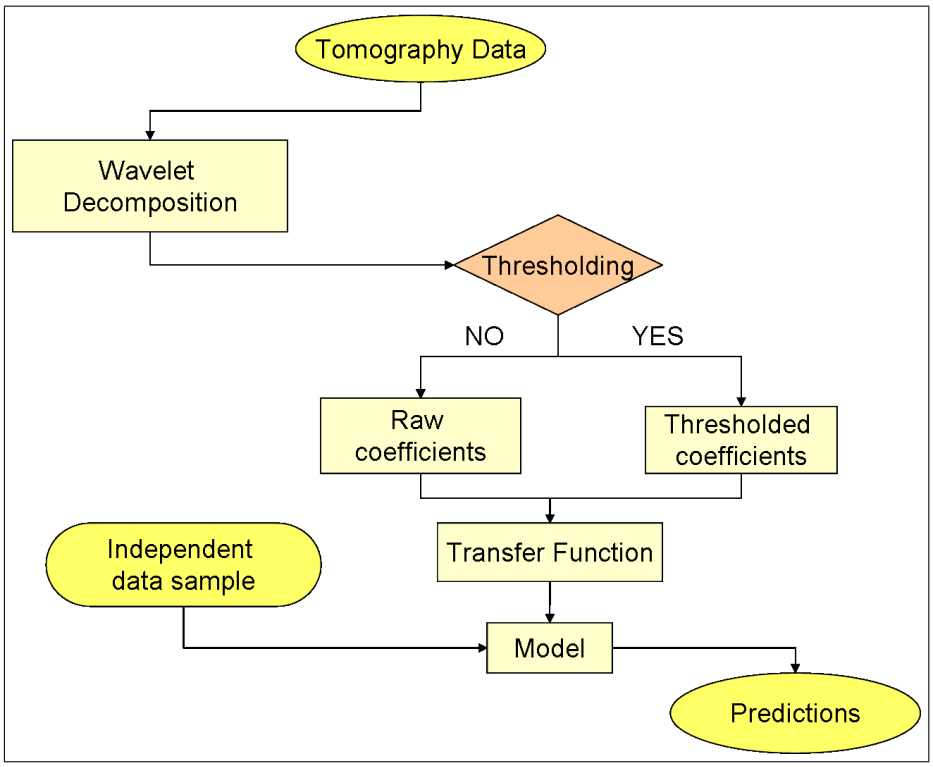


Figure 4.4: Modelling process.

The wavelet decomposition plots (Figure 4.3) show that the pattern of the wavelet coefficients is different at each resolution level for each flow regime. The fitted model will then aim to relate this different activity to the individual flow types.

4.3.2 Transfer function choice

The coefficients require some pre-processing in order to exploit the desired properties of the data. This is achieved by applying a *transfer function* to the wavelet coefficients before the model building stage.

In this problem there are 49 sets of wavelet coefficients at each time point, all representing slightly different regions of the pipe cross section. As the aim of the model will be to predict flow regime in the pipe at any given time, it would be sensible to combine the 49 coefficients in some sense due to the symmetry within the simulated data. Also, since we are interested in the level of activity (or frequency of the recorded signal) at each time point, it can be argued that the magnitude of the wavelet coefficients is what we are really interested in. This will avoid the possibility that large negative and large positive coefficients could be represented by a number close to zero after the averaging step. It is suggested the combining by taking the average of the absolute values of the 49 coefficients at each resolution level will give a reasonable measure.

If the wavelet coefficients of the ith electrode pair signal are represented by $\{d^i_{jk}\}$, where $i = 1, \ldots, 49$, and $\{d{*^i_{jk}}^*\}$ represents the thresholded coefficients, then the *activity measures* $\{a_{jk}\}$ used in the model building stage are defined as

No thresholding

$$a_{jk} = \frac{1}{49} \sum_i |d^i_{jk}| \tag{4.4}$$

With thresholding

$$a_{jk} = \frac{1}{49} \sum_i |{d^i_{jk}}^*| \tag{4.5}$$

where $j = 0, \ldots, J-1$, for $J = \log_2(n)$ and $k = 1, \ldots, n-1$, where n is the number of time points observed.

The thresholding method used here was the universal threshold of Donoho and Johnstone (1994). The aim was to see if thresholding could give any improvement in the model

predictions. Further refinements could be to consider different thresholding policies to optimise the model performance.

Other possible activity measures

There are several other ways in which the coefficients could be transformed to obtain an activity measure. One possibility is using the coefficient of variation, which is a normalised measure of dispersion and in this setting could be found over the 49 detail coefficients at each resolution level and location, again resulting in an 8 dimensional vector measure for each time point. Let $sd(.)$ denote the standard deviation and $\mu(.)$ denote the mean, then the coefficient of variation, c_{var}, is defined as $c_{var} = sd/\mu$. An alternative activity measure, η_{jk}, could then be calculated as

$$\eta_{jk} = \frac{sd(d^1_{jk}, \ldots, d^4_{jk}9)}{\mu(d^1_{jk}, \ldots, d^4_{jk}9)}.$$

Another similar possibility is the coefficient of dispersion, which is defined as the maximum absolute deviation divided by the median. This would result in an alternative activity measure, κ_{jk}, which could be defined as

$$\kappa_{jk} = \frac{1}{49} \sum_i \left| \frac{d^i_{jk} - \text{median}(d^i_{jk})}{\text{median}(d^i_{jk})} \right|.$$

4.3.3 Logistic modelling

Since the response variable is dichotomous, as two extreme flow regimes are being considered, logistic regression is used. The form of the model is

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_l \tag{4.6}$$

where p is the probability that the response is bubble flow, and hence $1-p$ the probability that the response is churn flow. The quantities $X_1, X_2, \ldots, X_l$ are the independent

predictor variables and the parameters $\beta_1, \beta_2, \ldots, \beta_l$, the regression coefficients, which have to be estimated from the data. The aim of building such a model is that given an observed signal, a prediction can be obtained from the transformed electrode-pair readings, giving the probability that the observed flow is from either a bubble or churn type flow regime. The logistic regression modelling idea extends to categorical variables with more than two values, using a multinomial approach. This methodology would allow the classification into more categories of flow regime (as in Section 4.3.5).

For training the logistic regression model, a dataset of 256 time points of churn flow, followed by 256 time points of bubble flow was used. Uncorrelated Gaussian noise, with constant standard deviation, was added to the signal. The non-decimated wavelet transform of the observed measurements, using the Haar wavelet, was found. The calculations were performed using the WaveThresh package (Nason, 2005) for the statistical programming environment R (R Development Core Team, 2007). For each electrode pair this produced nine time series of length 512, each representing the signal at differing frequencies.

At each time point, the activity measure of the wavelet coefficients was found, either before thresholding (equation 4.4) or after thresholding (equation 4.5), for each resolution level and time point. As the actual flow regime was known, this data could then be used to build a logistic regression model. In the model build the calculated activity measures $\mathbf{a_j} = (a_{j1}, \ldots, a_{jk}$ or $\mathbf{a_j}^* = (a^*_{j1}, \ldots, a^*_{jk})$ take the role of the $X_1, \ldots, X_l$ in equation 4.6, with $l = 1, \ldots, J$.

The model was used to predict flow regime in a set of independently generated flow patterns. This consisted of 168 time points of bubble flow, followed by 256 time points of churn flow, followed by 88 time points of bubble, with added Gaussian white noise with the same standard deviation as in the training dataset. We emphasise that the structure of the test dataset was different to that of the training dataset. The purpose of this was to demonstrate that the model predictions were not seriously affected by the structure of

the underlying dataset. The predictions generated by the model are probabilities of the time point being in bubble type flow. If these predicted probabilities were greater than or equal to 0.5, the model was regarded as predicting bubble flow, otherwise the prediction was churn flow. The correct classification rates (i.e. the percentage of time points which are classified correctly) for the models built using both thresholded and non-thresholded activity measures are shown in Figure 4.5.

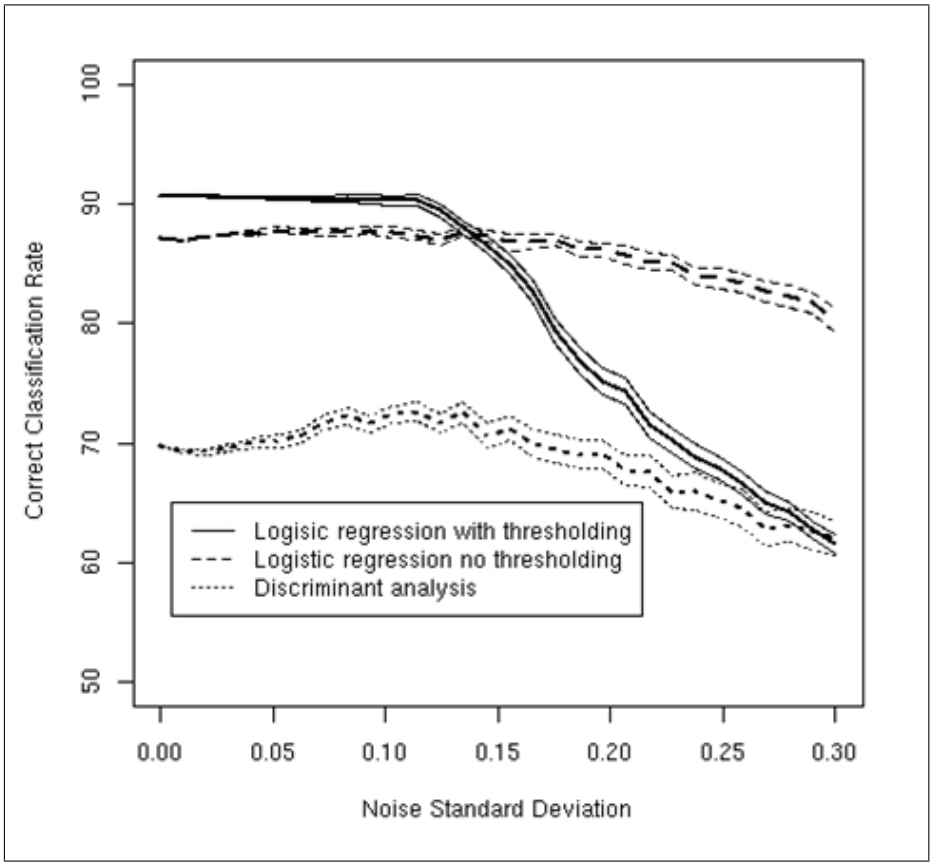


Figure 4.5: Correct classification rates for the logistic model for two regime model using thresholded, non-thresholded coefficients compared to a linear discriminant method based on the wavelet coefficients.

The error bands shown are ± 2 standard errors of the prediction rate over 100 replicates at each noise level.

The noise standard deviation is related to the signal to noise ratio by the formula

$$SNR = \frac{\text{signal standard deviation}}{\text{noise standard deviation}}.$$

The thresholded wavelet coefficients work better for noise levels below those corresponding to a SNR of 1. In this case the thresholded version is greatly superior, with the predictions being much more clearly separated into bubble and churn regimes. Beyond this the non-thresholded coefficients perform better, as the SNR decreases. The reason for this is that at low noise levels, thresholding is able to denoise the coefficients, whilst at higher noise levels, the signal will be swamped by noise and so the thresholding process will also damage the signal component.

For automated use of this procedure, it would be valuable to be able to determine from the data whether we are in a 'low noise' (high SNR) or 'high noise' (low SNR) situation. We could then use the thresholded or non-thresholded approach accordingly. Since the wavelet thresholding procedure involves estimation of the noise standard deviation, and the thresholded signal estimate can be used to estimate the signal standard deviation, this should be a feasible approach to producing an adaptive monitoring procedure.

4.3.4 Discriminant Analysis

In order to benchmark the performance of the model built in section 4.3.3, the alternative method of discriminant analysis was used to predict flow regime from the activity measures.

Discriminant analysis is an alternative to logistic regression, which has some prerequisite assumptions (independent variables need to be normally distributed, linearly related, or have equal within-group variances). However, discriminant analysis is preferred when the assumptions of linear regression are met since it has more statistical power than logistic regression.

Here the setup is that we have two known groups and each new observation must belong to either of these reference groups. The way discriminant analysis works is to assign the observation to the nearest group based on some distance measure. In this example, the Mahalanobis distance is used. Let $\bar{x}_i$ denote the reference group means, with $i \in (1, 2)$. In this situation, the distance measure of an observed point u from group i, $D^2(i)$ is given by:

$$D^2(i) = (u - \bar{x}_i)' \mathbf{W}^- 1 (u - \bar{x}_i)$$

, where $\mathbf{W}$ is the within groups variance of the two reference groups. This measure allows the two clusters of points to be assigned with some allowance for the within cluster variance. See Gnanadesikan (1997) for more details about discriminant analysis.

The activity measure used in the logistic regression modelling in the previous section resulted in a nine dimensional response for each time point, each number representing activity at a different frequency. It was anticipated that these activity vectors could be analysed using the statistical method of linear discriminant analysis (see for example Manly, 2005), which aims to map objects into one of several groups by means of their features and measurements. The training dataset from Section 4.3.3 was decomposed using the Daubechies Least Asymmetric (8) wavelet to generate activity measures as described above. A linear discriminant rule was then trained on these activity measures; this rule was then used to classify each time point in the test dataset to either bubble or churn flow.

Initially for the logistic regression models, Haar wavelets were used. This was motivated by their short support and therefore increased ability to detect high frequency activity. However, when using discriminant analysis, this basis performed poorly. It appeared that this was a result of the low correlation between the detail coefficients, which are then used in estimating the covariance matrix in order to establish a distance metric for the discriminant analysis. For this reason, a longer filter was opted for and so the Least Asymmetric with 8 vanishing moments was used. The longer filter means the

correlations are more stable and hence less prone to erratic behaviour due to a few outlying coefficients. This also suggests that a discriminant approach may be more sensitive to the choice of wavelet.

4.3.5 Multinomial logistic regression

We now consider data with three types of flow regime. The bubble and churn regimes from the previous section and an intermediate flow which lies somewhere between the two in terms of activity frequency. A multinomial logit model is fitted here as there are 3 response categories, that is bubble (state 1), churn (state 2) and intermediate (state 3). The modelling process used in R to fit a multinomial logit employs a neural network approach to estimate the probabilities p_{kl} that the kth observation is from state l, using the activity measures $\{a_{jk}\}$ as predictor variables (Venables and Ripley, 2002). To ensure that the resulting probabilities sum to 1, one category is set as the baseline (here chosen to be state 1), so that $p_{k1} = 1 - \sum_{l=2}^{3} p_{kl}$ and then we have

$$p_{kl} = \frac{\exp(\boldsymbol{a}_k^T \boldsymbol{\beta}_l)}{1 + \exp(\boldsymbol{a}_k^T \boldsymbol{\beta}_2) + \exp(\boldsymbol{a}_k^T \boldsymbol{\beta}_3)}, \qquad l = 1, 2, 3; \;\; k = 0, \ldots, n-1.$$

where $\boldsymbol{a}_k = (a_{0k}, \ldots, a_{(J-1)k})$ and the $\boldsymbol{\beta}$'s are vector valued regression coefficients. Further details of the multinomial model and the fitting algorithms used can be found in Faraway (2005).

A dataset composing of approximately equal proportions of the three flow regimes was used and white noise was added to the signal, representing measurement error. The non-decimated wavelet transform of the observed measurements, using the Haar wavelet, was found. For each electrode pair this produced nine activity level time series each of length 512, each representing the signal at differing frequencies. As the actual flow regime was known, this data could then be used to build a multinomial regression model. This model

was used to predict flow regime in a set of independently generated flow patterns, with similar proportions of each flow regime to the training data.

4.3.6 Correct classification rates

Figure 4.6 shows the correct classification rates for the models built using both thresholded and non-thresholded activity measures, averaged over 100 replications. Also shown are approximate point-wise 95% confidence intervals calculated using the standard errors from the replication.

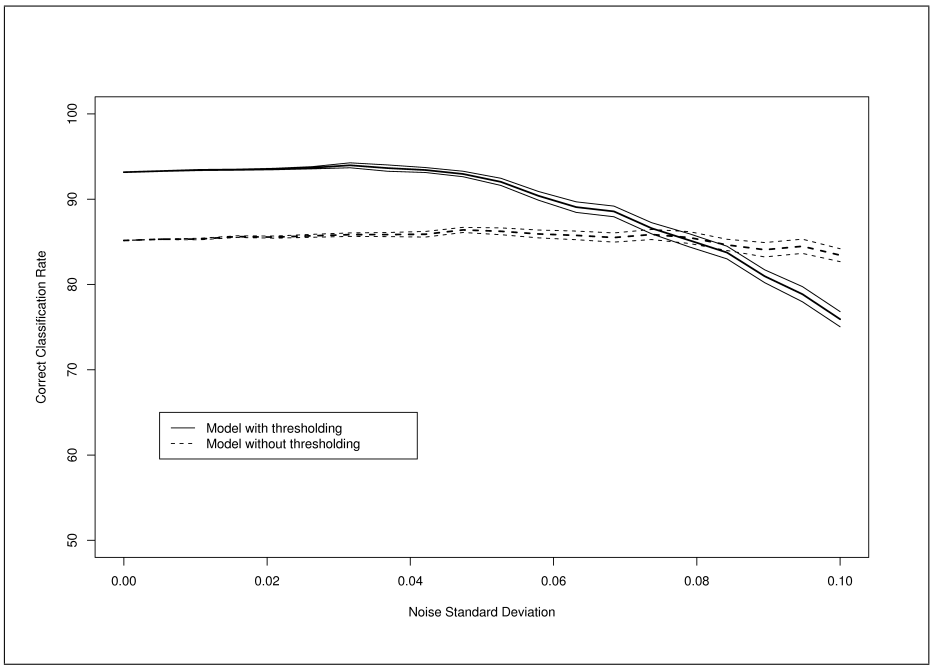


Figure 4.6: The correct classification rates for the model predicting three flow regimes, both with and without thresholding.

For all noise standard deviations below 0.07, it can be seen that the thresholded version outperforms the model without thresholding. Here, the performance of each method is approximately constant. The correct classification rate for the thresholded logistic

regression was approximately 93% for all noise standard deviations below 0.07. A standard deviation of 0.07 corresponds to a signal to noise ratio of 1; i.e. the "strength" of the signal and noise are roughly equal. For larger noise standard deviations the performance of the method with thresholding degrades, but the non-thresholded method appears to have a slower decrease in success rate.

4.3.7 Misclassification

The misclassification matrix resulting from the model is shown in Table 4.1. It can be seen the model successfully classifies at least 88% of all data points regardless of what flow regime it is actually from. This also shows that the main source of classification error is in classifying bubble flow as intermediate flow, which seems reasonable due to the similarity of the recorded data observed in Figure 4.2.

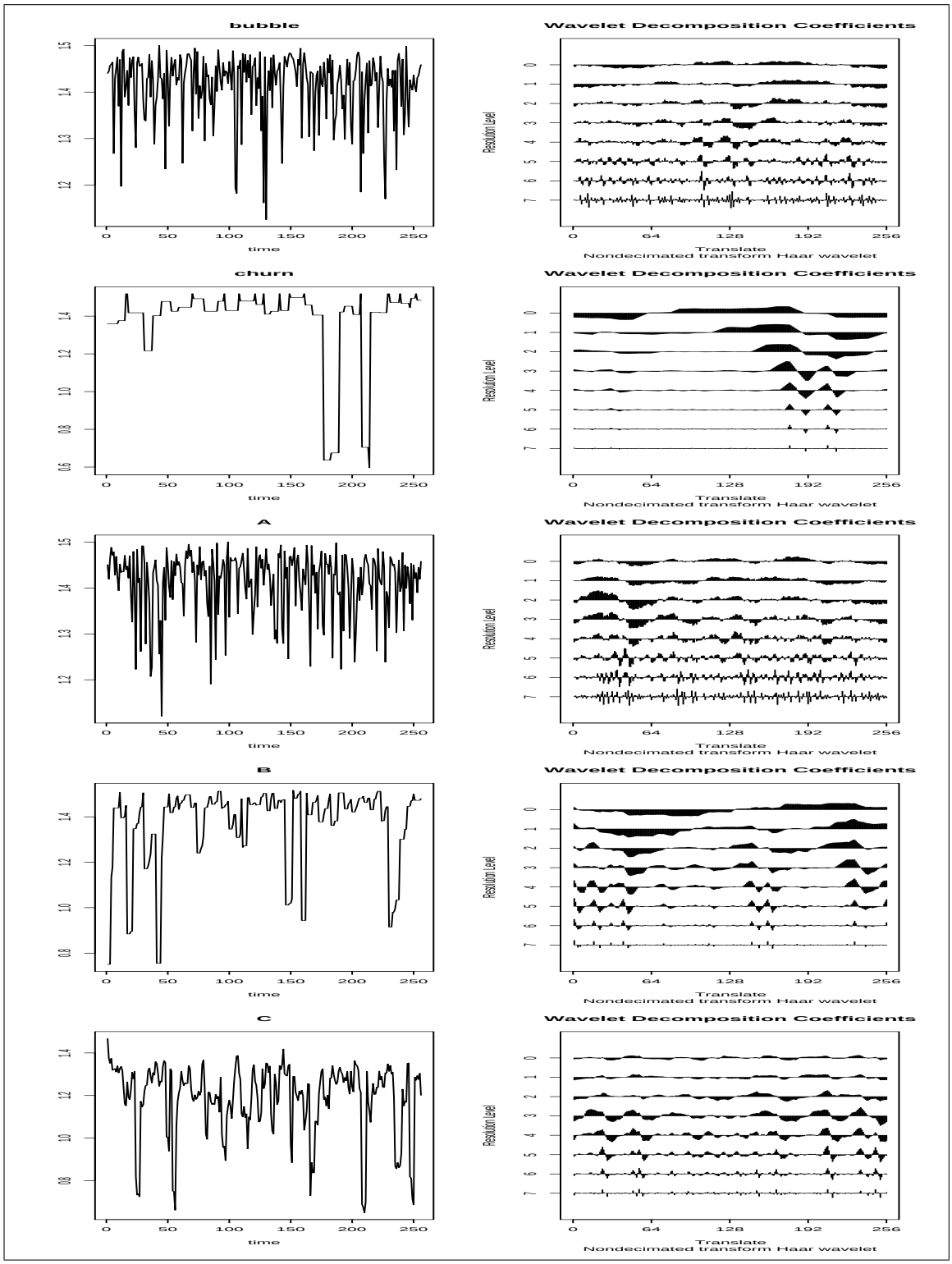


Figure 4.7: Some exploratory plots of the five flow regimes.

		True Flow Regime		
		bubble	churn	intermediate
Predicted	bubble	0.88	0.00	0.04
Flow	churn	0.04	0.99	0.00
Regime	intermediate	0.08	0.01	0.96

Table 4.1: Matrix of the correct cross-classification rates between the three flow regimes for a thresholded model with the added white noise standard deviation equal to 0.05.

4.4 Extending to more simulated datasets

4.4.1 Introduction

As in Section 4.3.5, it was possible to generate further datasets by varying the input parameters of bubble size and bubble length and considering more intermediate values between the bubble and churn extremes. This was done to obtain three further intermediate datasets (A, B and C). The relation to the original 'bubble' and 'churn' data was assumed to be unknown. A plot of an example of each of the five flow regimes used and the corresponding wavelet decomposition plots are shown in Figure 4.7. It was of interest to investigate how similar the three intermediate datasets were to the original. It was thought that the technique of multidimensional scaling would prove useful for this purpose.

4.4.2 Multidimensional scaling

Multidimensional scaling is defined as a collection of methods with 'the goal of detecting meaningful underlying dimensions that allow the researcher to explain observed similarities or dissimilarities between the investigated objects' (StatSoft, Inc., 2007).

Given a set of n points in Euclidean p-space it is possible to compute the distance between any pair of points and can obtain an $n \times n$ distance matrix or dissimilarity matrix (D). Multidimensional scaling answers the question of whether when given an $n \times n$ symmetrical matrix ($\Delta = \delta_{ij}$) of dissimilarities, can a configuration of points be found in Euclidean p-space (p open to choice) such that the calculated distance matrix (D) reasonably matches the given dissimilarity matrix (Δ).

Figure 4.8 shows how different multidimensional scaling techniques separate the flow regimes. It can be seen that there is no clear separation of the regimes using these

		True Flow Regime				
		Bubble	A	B	C	Churn
	Bubble	100%	47%	17%	39%	4%
Predicted	A	47%	100%	40%	39%	5%
Flow	B	17%	40%	100%	49%	14%
Regime	C	39%	39%	49%	100%	23%
	Churn	4%	5%	14%	23%	100%

Table 4.2: Similarity matrix (thresholded) of the five simulated datasets.

techniques.

4.4.3 Investigating the data using logistic regression

For a measure of similarity between each of the 5 simulated datasets, each pairwise combination was analysed in a similar fashion to the bubble and churn datasets in Section 4.3.3. This produced a correct classification probability for each pairwise combination of datasets, which was then used as a measure of similarity.

The similarity matrix obtained for the thresholded model was found and is shown in Table 4.2. The similarity for the non-thresholded models is also shown in Table 4.3.

These similarities will be used later (Section 4.5) in assessing how well a model performs dependant upon how well it classified flow into ‘similar’ datasets.

Comparing Tables 4.2 and 4.3 shows that when the underlying data has been thresholded, the similarity of each of the flow regimes increases. This is intuitive since the differences between some of the higher frequency regimes may have been observable at finer resolution levels, information which may be removed by thresholding. It can be seen that regime A is most similar to bubble flow. For the thresholded version, regime C has

		True Flow Regime				
		Bubble	A	B	C	Churn
	Bubble	100%	37%	7%	3%	7%
Predicted	A	37%	100%	27%	5%	3%
Flow	B	7%	27%	100%	36%	32%
Regime	C	3%	5%	36%	100%	16%
	Churn	7%	3%	32%	16%	100%

Table 4.3: Similarity matrix (not thresholded) of the five simulated datasets.

relatively high similarities to all of the other regimes, suggesting that this flow combines characteristics of both bubble and churn flows. Regime B is most like regime C, but also has similarities to churn flow. This may mean that differentiating between B and C will be difficult.

If we consider the dissimilarity matrices, derived from Tables 4.2 and 4.3 using the relationship *dissimilarity = 100% - similarity*, as distance matrices, multidimensional scaling can be used to obtain a visual representation of the similarities of the flow regimes. Figure 4.9 shows the result of classical scaling, although Kruskal and Sammon methods gave similar results.

From Figure 4.9 it can be seen that for the non-thresholded models, the first dimension separates the flows into two groups, bubble with A and the churn with B and C. The second scaling dimension then separates A from bubble and B and C from churn, but the separation between B and C in this second dimension is not as great. For the thresholded version, flows B and C are very similar in the first dimension.

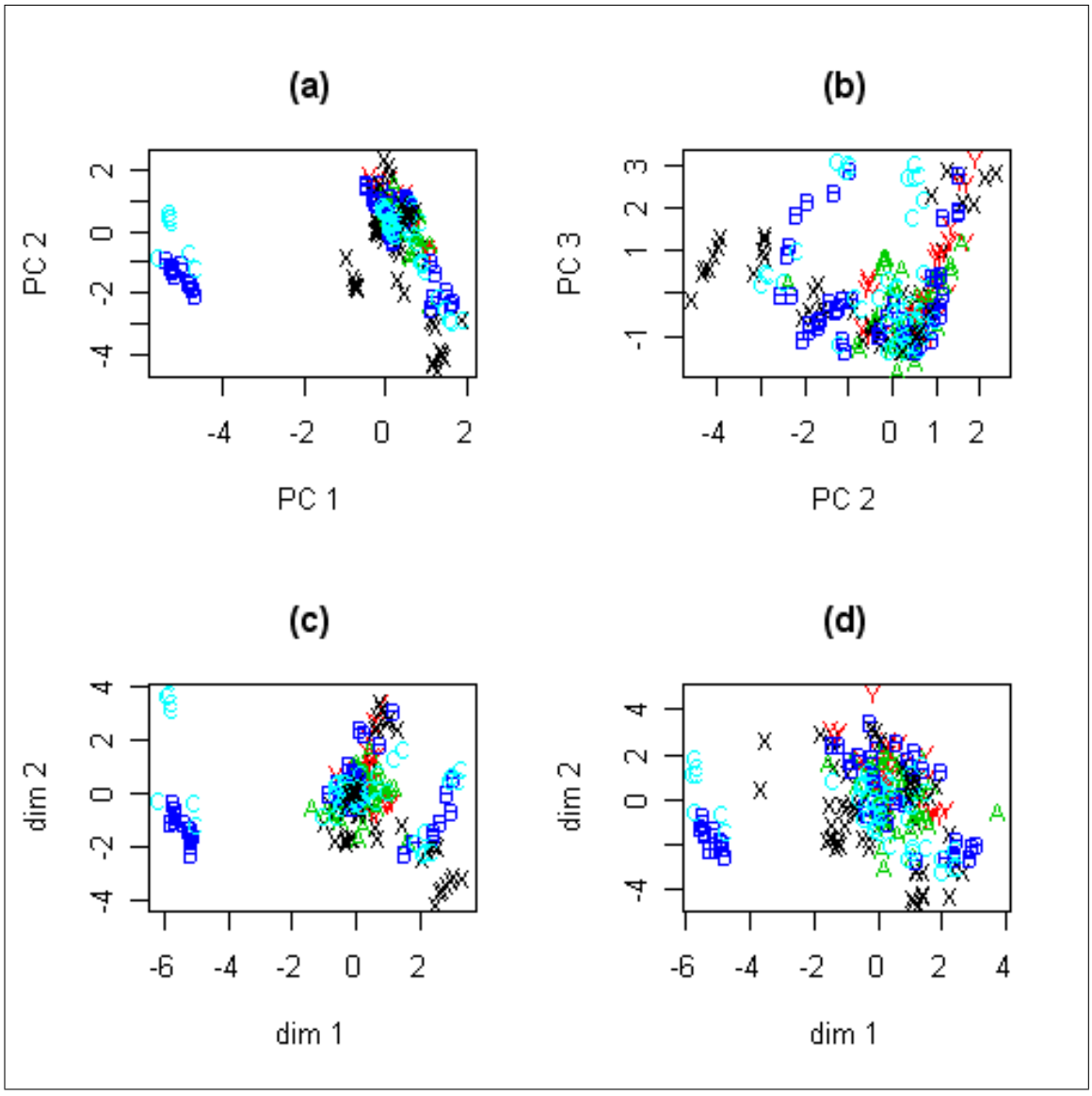


Figure 4.8: Multidimensional scaling of the recorded data. (a) and (b) shows plots of the 1st, 2nd and 3rd principal components, (c) shows the first two dimensions using Sammon mapping and (d) shows the first two dimensions using Kruskal mapping.

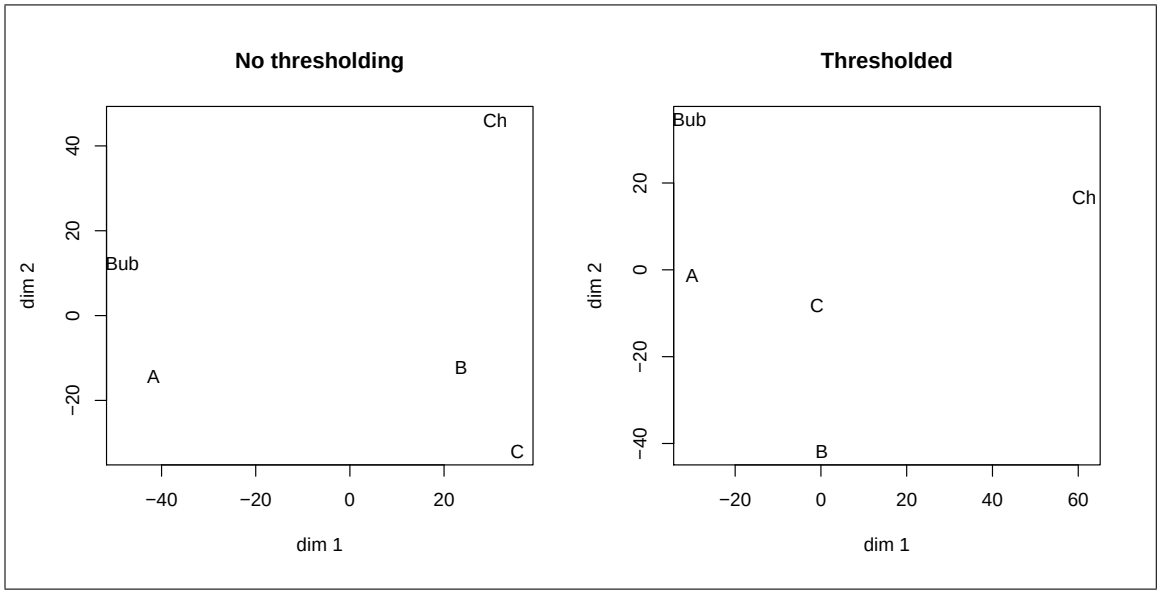


Figure 4.9: Using Multidimensional scaling to separate the flow regimes.

4.4.4 Alternative use of MDS to separate flow regimes

An alternative measure of distance between the flow regimes could be considered. This time, the activity measures were found using equation 4.4 for each of the test datasets. This produced 256 time points represented at each of 8 resolution levels for each flow regime. The mean of the activity level

$$\bar{a_k} = \frac{1}{\sum j} \sum_j a_{jk}$$

was then found at each time point. Each flow regime then represented by an 8-vector. The reason for using the mean was that the flow regime will be represented over all 256 time points and so a typical activity level might help to identify differences between flow regimes in terms of the activity level representation.

The Euclidean distance between the 8-vectors was found and used in Kruskal's non-metric scaling method (Cox and Cox, 2001) to map the points into 4 dimensions. Kruskal's scaling method minimises the 'stress' of the configuration, S, given by:

$$S = \left(\frac{\sum_{j<k} (\theta(a_{jk}) - \hat{a}_{jk})^2}{\sum_{j<k} a_{jk}^2} \right)^{\frac{1}{2}}$$

. This resulted in a solution which was more tightly clustered and a representation of the fitted configuration is shown in Figure 4.10.

4.5 Further multinomial examples

It was also possible to fit a multinomial model to a dataset consisting of a compilation of four or five of the flow regimes. The purpose of this was to check for model consistency and explore how well the modelling technique would perform when the test data are more similar in terms of their frequency characteristics.

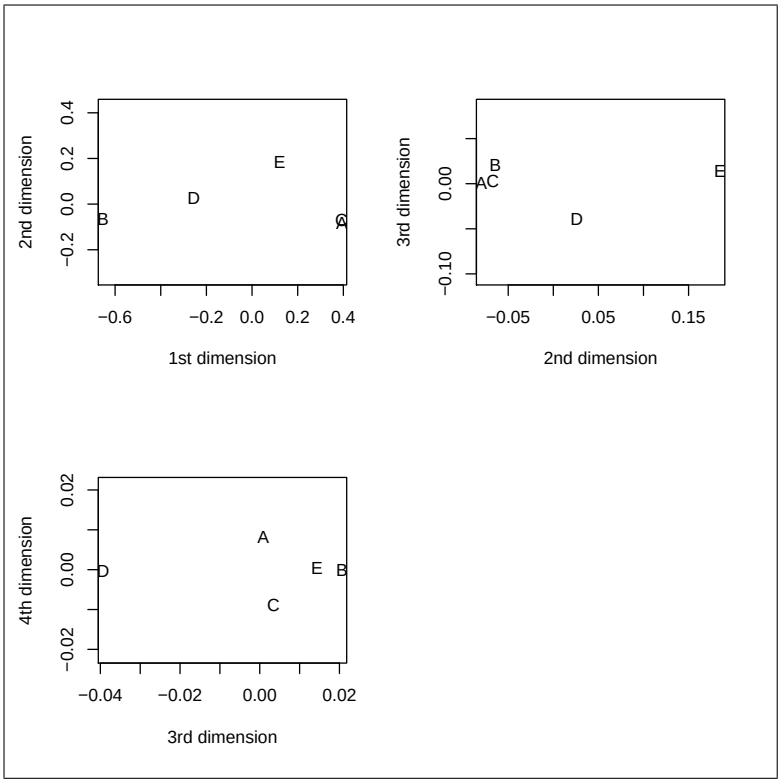


Figure 4.10: Kruskal scaling applied to the activity level measures. Key: A — Bubble, B — Churn, C — Flow A, D — Flow B, E — Flow C.

It can be seen from Figure 4.7 that the intermediate flows A, B and C are more visually similar than the two extremes of bubble and churn flow. This can also be seen from the multidimensional scaling approach in Figure 4.9. To test the performance of the method, first a 4 level response multinomial model was fitted to a regime consisting of a combination of the bubble, churn, flow A and flow B. As a further test the method was then tried on all 5 flow regimes.

The 4 flow example consisted of 256 time points of each flow in the build sample. The test sample consisted again of 256 blocks, but with a alternative ordering of the flow regimes.

The 5 flow regime example consisted of 205 time points of each of the first four flows, together with 204 time points of the final intermediate flow. This order was again changed

		True Flow Regime			
		Bubble	Churn	A	B
Predicted	Bubble	36%	13%	71%	6%
Flow	Churn	4%	64%	15%	33%
Regime	A	18%	13%	14%	6%
	B	42%	11%	0%	55%

Table 4.4: Correct classification rates for the 4 flow regime model.

in the model test sample.

4.5.1 Results

Multinomial model with 4 response levels

As shown in Table 4.4, the model performed best on the lower frequency regimes, with a Churn correct prediction rate of 64% and that of Flow B being 55%. As Flow B is most similar to Churn, if classifying as either of these regimes is considered a partial success then the classification rates increase to 75% and 88% respectively.

The worst correct classification was that of Flow A, which was chronically predicted to be Bubble flow. However, it is clear from Figure 4.7 that these two regimes do share similarities. A prediction rate of 85% is observed when considering whether Flow A was classified as a high frequency flow regime. The model misclassified 64% of the bubble flow as other flow types, yet most of the wrong predictions fell into the Flow A category.

The addition of a further flow regimes has, in this case, decreased the reliability of the model. Despite this, the model is still able to classify on the broader scales of high or low frequency regime.

		True Flow Regime				
		Bubble	Churn	A	B	C
	Bubble	100%	5%	1%	28%	0%
Predicted	Churn	0%	5%	0%	28%	100%
Flow	A	0%	0%	28%	0%	0%
Regime	B	0%	90%	70%	43%	0%
	Churn	0%	0%	0%	0%	0%

Table 4.5: Correct classification rates for the 5 flow regime model.

Multinomial model with 5 response levels

Looking at Table 4.5, it is apparent that the modelling method breaks down further as one more possible regime is added. The model predicts 100% of Bubble flow correctly, yet is unstable for all of the other regimes. The modelling method seems unable to predict with any reliability when 5 flows are considered.

4.6 Conclusions

Wavelets have been shown to be a useful tool in extracting information from time series with changing frequencies. This chapter has identified some methods that can be used in an electrical tomographic setting, where such frequency changes exist.

The *transfer function* was identified as an important step in the modelling process. This encoded the desired characteristics of the data to allow interpretable models to be built.

Logistic and multinomial regression techniques can be used to predict flow regime, with varying degrees of success. For distinguishing between up to 3 flow regimes the modelling method performs well. When the demands of more regimes was introduced, the method

was no longer able to accurately predict.

Various methods exist in the literature for the analysis of tomographic data, including extended Kalman filtering (Ijaz et al., 2006) and the modified Newton Raphson method (Williams and Xie, 1993). However, the applications vary and there is no direct comparison of the performance of the techniques developed to these existing methodologies.

Chapter 5

Multiwavelet modelling approaches

5.1 Introduction

5.1.1 What are multiwavelets?

Scalar wavelets as described in Section 2.4 are based on one father wavelet or scaling function. Wavelet families with more than one scaling function are known as multiwavelets and Strang and Strela (1995) described several reasons why multiwavelets offer advantages over their scalar counterparts. A scalar wavelet system cannot combine symmetry, orthogonality and higher order approximation as is possible with a multiwavelet system.

The theory of multiwavelets is based on the idea of multiresolution analysis (MRA), similar to that of scalar wavelets (see Section 2.3). However, the basis for V_0 is generated by translates of N scaling functions $\phi_1(t-k), \ldots, \phi_N(t-k)$. The vector $\Phi(t) = [\phi_1(t), \ldots, \phi_N(t)]^T$, will satisfy a matrix dilation equation

$$\Phi(t) \;=\; \sum_k \mathbf{H_k} \Phi(2t-k). \tag{5.1}$$

The coefficients $\mathbf{H_k}$ are N by N matrices. A similar relationship holds for the mother wavelets $\psi_1(t), \ldots, \psi_N(t)$, which satisfy the matrix equation

$$\Psi(t) \quad = \quad \sum_k \mathbf{G_k} \Phi(2t - k). \tag{5.2}$$

Again, here the coefficients $\mathbf{G_k}$ are N by N matrices.

Equations 5.1 and 5.2 are analogous to the dilation equations in the scalar setting. Strang and Strela (1995) gives a discussion on how these relations can be reached by considering wavelet theory as the 'iterated limit of filter bank theory'.

5.1.2 Prefiltering

The input of a multiwavelet transform must be vector valued. To achieve this when applying multiwavelets to scalar series, the input must initially be preprocessed using a prefilter. There are a variety of prefilters in the literature designed to produce vector inputs, allowing scalar series to be analysed using multiwavelet methods. Some examples of commonly used prefilters are:

- **Identity** — this is when a series $x_1, \ldots, x_n$ is 'stacked' to create a series of vectors, so for a 2 dimensional multiwavelet transform this series would become
$$\left\{ \begin{bmatrix} x_1 \\ x_1 \end{bmatrix}, \ldots, \begin{bmatrix} x_n \\ x_n \end{bmatrix} \right\}$$
.

- **Repeated row** — For this prefilter, the series x_n would become
$$\left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \ldots, \begin{bmatrix} x_{n-1} \\ x_n \end{bmatrix} \right\}$$
.

Strela et al. (1999) propose using a multiwavelet transform for images, where the input series are then themselves 2-dimensional, removing the need for prefiltering.

5.1.3 Filter choice

There are relatively few examples of the implementation of different multiwavelet filters compared to scalar wavelet. These include the 2-dimensional Geronimo-Donovan-Hardin-Massopust (GDHM) and Chui-Lian families and the 3-dimensional Donovan family. Higher dimensional multiwavelets exist, but are difficult to obtain explicitly due to the high dimensional orthogonality conditions.

In the case of a scalar input series, the choice of dimension for the multiwavelet family and prefilter would be based on choice of support, symmetry or the number of vanishing moments as in the scalar wavelet case.

Some of the properties motivating the choice of the GDHM wavelet filter will be discussed in the next section.

5.1.4 Example of a multiwavelet system

One of the simplest multiwavelets bases is that constructed by Geronimo et al. (1996). This consists of two scaling functions $\phi_1(t), \phi_2(t)$, together with two associated mother wavelets $\psi_1(t), \psi_2(t)$, these are shown in Figure 5.1.

This filter has some properties which are not able to be combined in the case of scalar wavelets, that is symmetry, orthogonality and second order approximation. It also has short support, with each of the associated scaling functions having support $[0, 2]$ and the wavelets having support $[0, 1]$ and $[0, 2]$. It can be seen in Figure 5.1 that both scaling functions are symmetric, and the wavelets form a symmetric/antisymmetric pair.

The dilation and wavelet equations for this system have four coefficients:

$$\begin{aligned} \mathbf{\Phi(t)} &= \begin{bmatrix} \phi_1(t) \\ \phi_2(t) \end{bmatrix} \\ &= \mathbf{H_0\Phi}(2t) + \mathbf{H_1\Phi}(2t-1) + \mathbf{H_2\Phi}(2t-2) + \mathbf{H_3\Phi}(2t-3) \end{aligned}$$

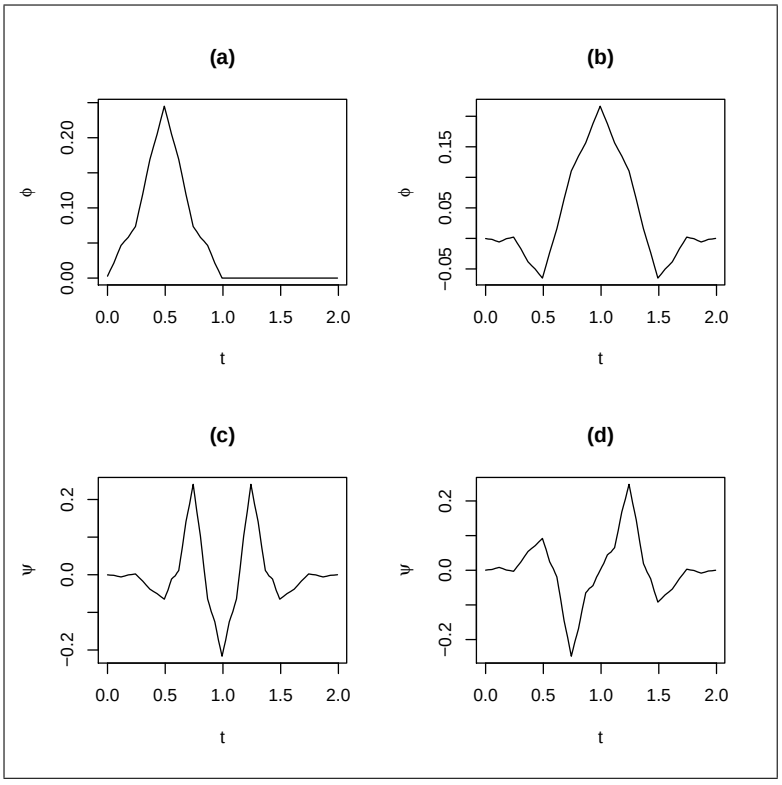


Figure 5.1: The GDHM multiwavelet basis functions (a) and (b) are the two scaling functions, with (c) and (d) being the two mother multiwavelets.

with

$$\mathbf{H_0} = \begin{bmatrix} \frac{3}{5\sqrt{2}} & \frac{4}{5} \\ -\frac{1}{20} & -\frac{3}{10\sqrt{2}} \end{bmatrix}, \quad \mathbf{H_1} = \begin{bmatrix} \frac{3}{5\sqrt{2}} & 0 \\ -\frac{9}{20} & \frac{1}{\sqrt{2}} \end{bmatrix},$$

$$\mathbf{H_2} = \begin{bmatrix} 0 & 0 \\ \frac{9}{20} & -\frac{3}{10\sqrt{2}} \end{bmatrix}, \quad \mathbf{H_3} = \begin{bmatrix} 0 & 0 \\ -\frac{1}{20} & 0 \end{bmatrix}.$$

$$\begin{aligned} \mathbf{\Psi(t)} &= \begin{bmatrix} \psi_1(t) \\ \psi_2(t) \end{bmatrix} \\ &= \mathbf{G_0\Phi}(2t) + \mathbf{G_1\Phi}(2t-1) + \mathbf{G_2\Phi}(2t-2) + \mathbf{G_3\Phi}(2t-3) \end{aligned}$$

$$\mathbf{G_0} = \begin{bmatrix} -\frac{1}{20} & -\frac{3}{10\sqrt{2}} \\ \frac{1}{10\sqrt{2}} & \frac{3}{10} \end{bmatrix}, \quad \mathbf{G_1} = \begin{bmatrix} \frac{9}{20} & -\frac{1}{\sqrt{2}} \\ -\frac{9}{10\sqrt{2}} & 0 \end{bmatrix},$$

$$\mathbf{G_2} = \begin{bmatrix} \frac{9}{20} & -\frac{3}{10\sqrt{2}} \\ \frac{9}{10\sqrt{2}} & -\frac{3}{10} \end{bmatrix}, \quad \mathbf{G_3} = \begin{bmatrix} -\frac{1}{20} & 0 \\ -\frac{1}{10\sqrt{2}} & 0 \end{bmatrix}.$$

5.2 The decimated multiwavelet transform computation

The decimated multiwavelet transform can be performed in an analogous way to scalar wavelets using matrix valued high and low pass filters, $\mathcal{G} = \{\mathbf{G_k}\}$ and $\mathcal{H} = \{\mathbf{H_k}\}$, where in the GDHM case $k = 1, \ldots, 4$.

Consider a scalar sequence of length 2^J, $x_1, \ldots, x_n$. Top obtain the wavelet decomposition using the GHDM filter as described above, the series would find be prefiltered to a sequence of 2-d vectors.

$$x_1, \ldots, x_n \xrightarrow{\text{prefilter}} \left\{ \begin{bmatrix} x_1 \\ x_1 \end{bmatrix}, \ldots, \begin{bmatrix} x_n \\ x_n \end{bmatrix} \right\}$$

The next step is to filter this sequence in a similar way to the scalar case using the high and low pass filters, $\mathcal{G}$ and $\mathcal{H}$ respectively. The main difference here is that the filtering step involves matrix multiplication of the vectors in the original sequence.

5.2.1 Example of the decimated MWT

For a dataset containing 4 vectors, c_j with $J = 2$, the multiwavelet transform can be computed as follows.

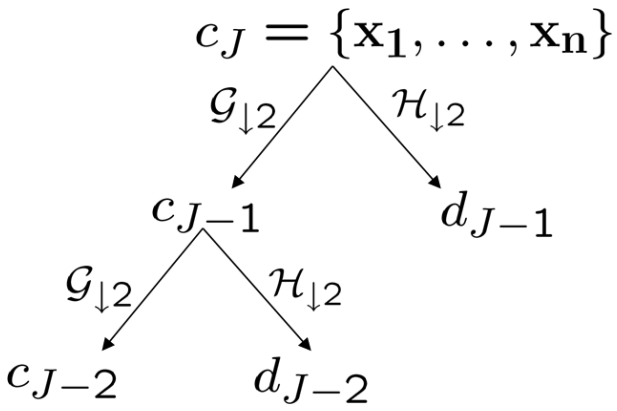


Figure 5.2: Diagram to show how the first two levels of the decimated multiwavelet decomposition would be found, $\downarrow 2$ represents the decimation step.

1. Take the initial dataset

$$\begin{aligned} \mathbf{c_2} &= \{\mathbf{c_{2,0}}, \mathbf{c_{2,1}}, \mathbf{c_{2,2}}, \mathbf{c_{2,3}}\} \\ &= \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 5 \\ 6 \end{bmatrix}, \begin{bmatrix} 7 \\ 8 \end{bmatrix} \right\}. \end{aligned}$$

2. Find the first level detail coefficients by applying the high pass filter $\mathcal{G}$ to $\mathbf{c_2}$

$$\begin{aligned} \mathbf{d_1} &= \mathcal{G}_{\downarrow 2}\mathbf{c_2} \\ &= \{(\mathbf{G_0 c_{2,0}} + \mathbf{G_1 c_{2,1}} + \mathbf{G_2 c_{2,2}} + \mathbf{G_3 c_{2,3}}), (\mathbf{G_0 c_{2,2}} + \mathbf{G_1 c_{2,3}} + \mathbf{G_2 c_{2,0}} + \mathbf{G_3 c_{2,1}})\} \\ &= \left\{ \begin{bmatrix} -1.325483 \\ -0.3514719 \end{bmatrix}, \begin{bmatrix} -4.153911 \\ -2.476955 \end{bmatrix} \right\} \end{aligned}$$

3. The next level of scaling coefficients can be found be applying the low pass filter

$\mathcal{H}$.

$$\begin{aligned}\mathbf{c_1} &= \mathcal{H}_{\downarrow 2}\mathbf{c_2} \\ &= \{(\mathbf{H_0c_{2,0}} + \mathbf{H_1c_{2,1}} + \mathbf{H_2c_{2,2}} + \mathbf{H_3c_{2,3}}), (\mathbf{H_0c_{2,2}} + \mathbf{H_1c_{2,3}} + \mathbf{H_2c_{2,0}} + \mathbf{H_3c_{2,1}})\} \\ &= \left\{ \begin{bmatrix} 3.297056 \\ 9.891169 \end{bmatrix}, \begin{bmatrix} 4.331371 \\ 7.159798 \end{bmatrix} \right\}\end{aligned}$$

4. Apply the low pass filter $\mathcal{H}$ to obtain the finest level of scaling coefficients.

$$\begin{aligned}\mathbf{c_0} &= \mathcal{H}_{\downarrow 2}\mathbf{c_1} \\ &= (\mathbf{H_0c_{1,0}} + \mathbf{H_1c_{1,1}} + \mathbf{H_2c_{1,0}} + \mathbf{H_3c_{1,1}}) \\ &= \begin{bmatrix} 9.060387 \\ 8.500387 \end{bmatrix}\end{aligned}$$

5. Again, the high pass filter $\mathcal{G}$ can be applied to obtain the finest level of multiwavelet coefficients.

$$\begin{aligned}\mathbf{d_0} &= \mathcal{G}_{\downarrow 2}\mathbf{c_1} \\ &= (\mathbf{G_0c_{1,0}} + \mathbf{G_1c_{1,1}} + \mathbf{G_2c_{1,0}} + \mathbf{G_3c_{1,1}}) \\ &= \begin{bmatrix} -1.625097 \\ -4.662742 \end{bmatrix}\end{aligned}$$

Then the DMWT of $\mathbf{c_2}$ will be $\{\mathbf{d_1}, \mathbf{d_0}, \mathbf{c_0}\}$.

5.3 Non-decimated multi-wavelet transform

The non-decimated multiwavelet transform (NDMWT) can be computed in a similar fashion to the decimated case. However, analogous to the scalar case (Section 2.8) the high and low pass filters are padded with alternate zero matrices at each level. These filters are then applied to the data, producing sequences of multiwavelet coefficients at each level that are of the same length and dimension as the input series.

Following on from the example in Section 5.2.1, the filters are padded at each level with a (2×2) zero matrix. The resulting wavelet coefficients are:

$$\mathbf{c_0} = \left\{ \begin{bmatrix} 9.060387 \\ 8.500387 \end{bmatrix}, \begin{bmatrix} 12.470580 \\ 3.677645 \end{bmatrix}, \begin{bmatrix} 11.323128 \\ 5.300387 \end{bmatrix}, \begin{bmatrix} 7.007838 \\ 11.403128 \end{bmatrix} \right\}$$

$$\mathbf{d_0} = \left\{ \begin{bmatrix} -1.625097 \\ -4.662742 \end{bmatrix}, \begin{bmatrix} 3.580589 \\ -2.262742 \end{bmatrix}, \begin{bmatrix} -0.8250967 \\ 4.6627417 \end{bmatrix}, \begin{bmatrix} 1.649218 \\ 2.262742 \end{bmatrix} \right\}$$

5.4 Modelling seismic data using multi-wavelets

5.4.1 Motivation

In Section 3.6 it was observed that incorporating neighbouring traces in the model build improved the reconstruction of the underlying truth. By considering sets of neighbouring traces at distinct points in time, it is possible to treat the data as a vector values series. This can then be investigated using a multiwavelet transform to yield an alternative set of wavelet coefficients, which could then be used to build a predictive model for the underlying acoustic impedance.

5.4.2 How to use the wavelet coefficients

As illustrated in Section 5.2.1, the wavelet coefficients obtained from the multiwavelet transform will themselves be vectors. This means that the *transfer function* (Section 3.4) will have to be altered to accommodate this.

In order to use the coefficients in the modelling approach, a few different transforms of the coefficients will be explored. For a p dimensional multiwavelet transform, giving

coefficients of the form $\mathbf{d}_{jk} = (d^1_{jk}, \ldots, d^p_{jk})^{\mathbf{T}}$, some possible ways to incorporate these into a modelling framework would be:

- **Modulus** — use $||d_{jk}|| = \left((d^1_{jk})^2 + \cdots + (d^p_{jk})^2\right)^{\frac{1}{2}}$.
- **Modulus-argument representation** The set of angles, $\{\theta_1, \ldots, \theta_{p-1}\}$, which define the vector could be used alongside the modulus to obtain a different representation of the coefficients.
- **Splitting** — consider each of the series $\{d^1_{jk}\}, \ldots, \{d^p_{jk}\}$ separately.
- **Scaling the coefficients** — the observed coefficients will have different variances in each component direction. It is possible to map these vector valued coefficients, $\mathbf{d}_{jk}$, into a scalar, θ_{jk}, taking account of these differing variances. This mapping will be of the form

  $$\theta_{jk} = \mathbf{d}^T_{jk} V_j^{-1} \mathbf{d}_{jk}$$

  where V_j is the resolution level covariance matrix, robustly estimated from the observed coefficients (Huber, 1981). This measure is also the Mahalanobis distance of the point $\mathbf{d_{jk}}$ from the origin. Barber and Nason (2004) use this sort of measure in the conext of complex wavelets.

5.4.3 Model framework

As the setup of the original dataset was that of adjacent data readings, no prefilter will be used, instead the observed traces will be grouped up into 2 or 3 adjacent traces, depending on the dimension of the filter used.

To model the underlying acoustic impedance trace, a linear regression approach will be used as in Section 3.3.1. When the idea of introducing neighbouring traces was introduced, one trace from either side was considered. This will be reproduced here,

using the 3 dimensional Donovan Wavelet, described in Donovan et al. (1996). Several different ways of incorporating the multiwavelet coefficients into the modelling process will be explored.

5.4.4 Multiwavelet coefficients

As with the scalar wavelet transform, the coefficients will give a decomposition of the underlying signal in both time and frequency. In order to identify properties of the coefficients which might explain characteristics of the underlying signal, it will be useful to look at the coefficients after the various transforms have been applied.

Split

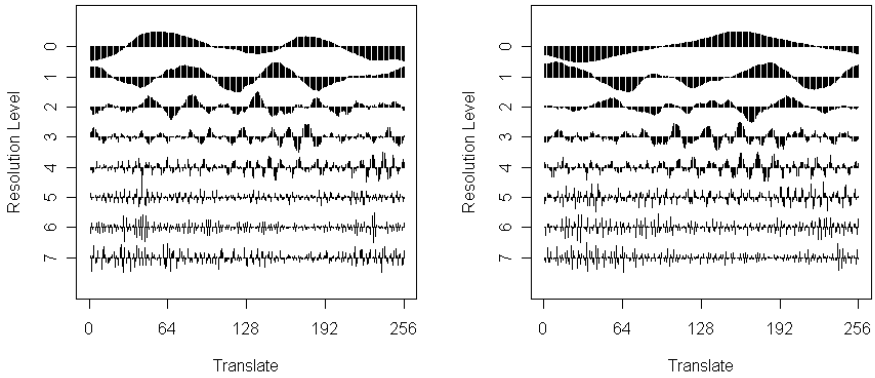
Figure 5.3 shows the multiwavelet decomposition of the data, using both the GDHM and Donovan filters. Each graph shows one dimension of the wavelet coefficients at each resolution level. For both filters, it can be seen that there is increased activity in resolution levels 3 and 4 in the central region of higher acoustic impedance. At the finest level, it appears as though more activity is observed outside this region, corresponding to the lower acoustic impedance level.

Modulus-argument representation

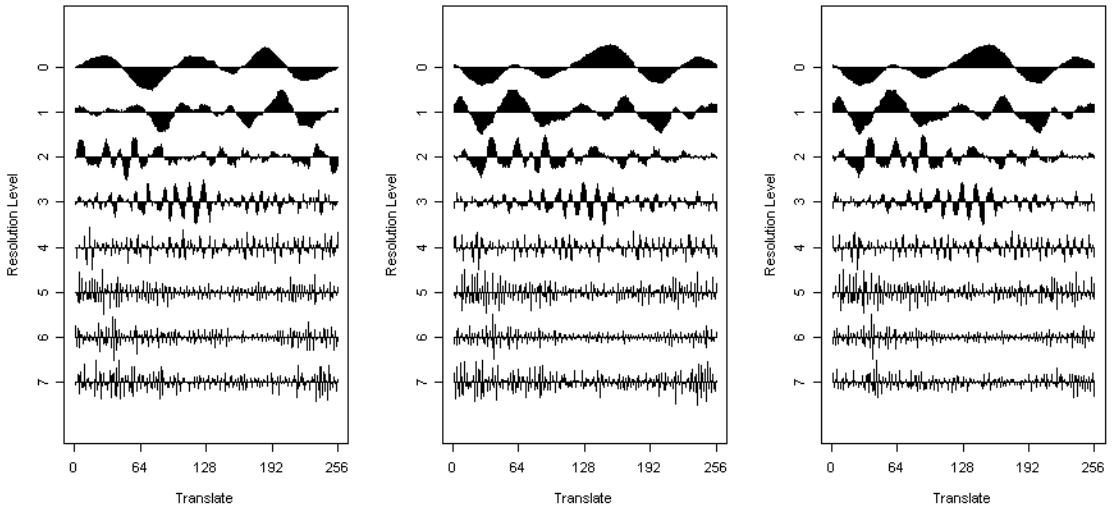
The multiwavelet decomposition of the data using both filters can again be seen in Figure 5.4. This time the left most graph shows the modulus of the coefficient. The following graphs show a representation of the angles of the vector, transformed to show their difference from a zero angle to allow comparison. The modulus-argument representation makes it difficult to pick out any regions of interest. This suggests that this transform of the coefficients is masking the properties which we are looking to extract.

Scaled

Robustly estimating the covariance matrix and transforming the $\mathbf{d_{jk}}$ to θ_{jk}'s gives a scalar representation of the coefficients, which is shown in Figure 5.5. The scaled coefficients show clumps of activity, yet these do not line up with where one would expect/ This again suggests that this transform will not be suitable for exploiting the frequency characteristics.

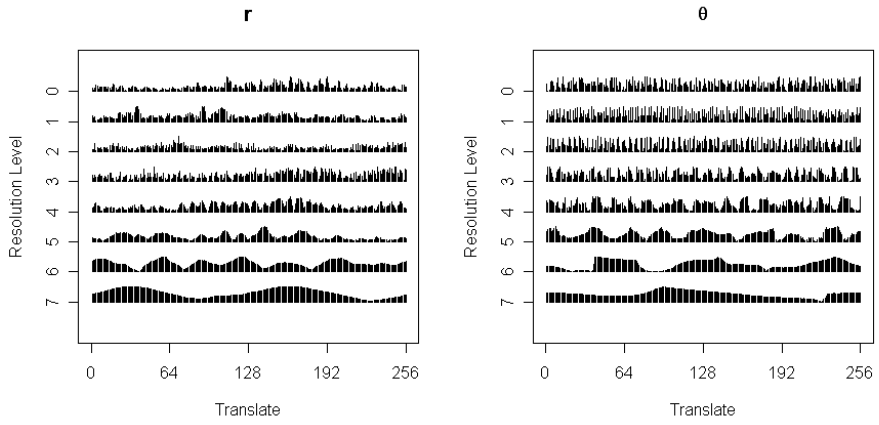


(a) Geronimo

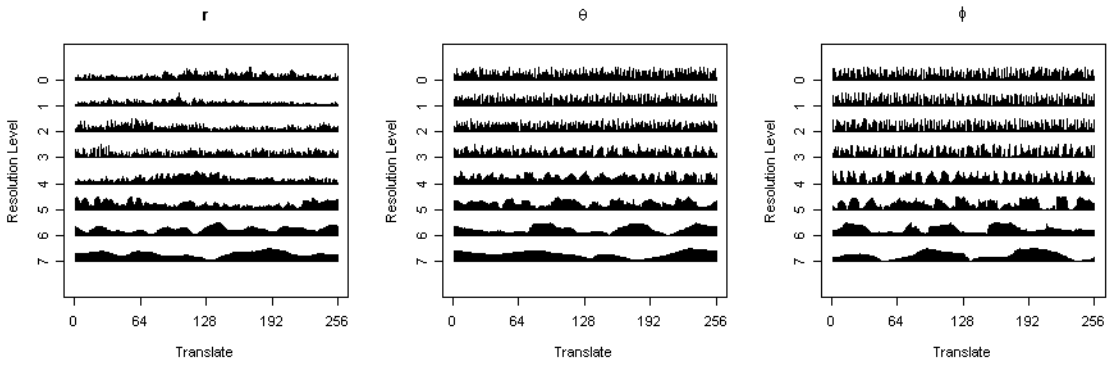


(b) Donovan

Figure 5.3: Split multiwavelet coefficients

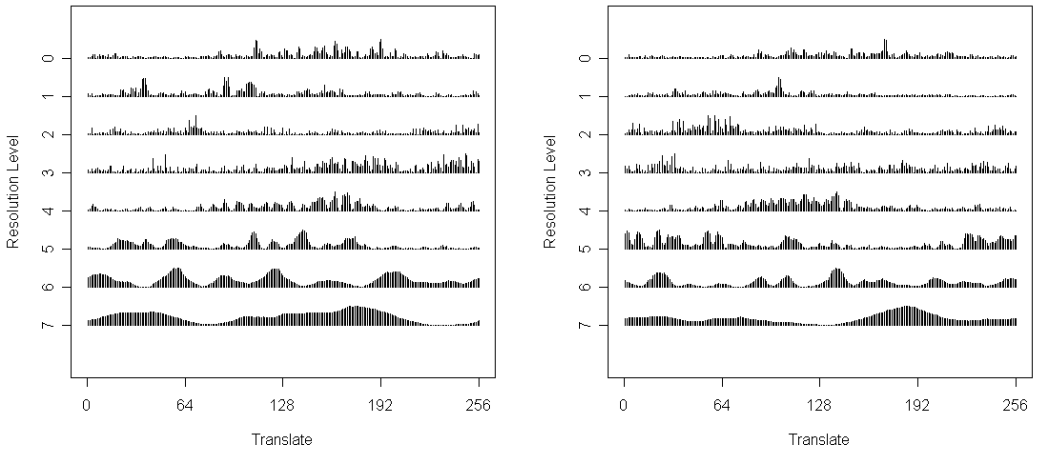


(a) Geronimo



(b) Donovan

Figure 5.4: Modulus-argument representation of the multiwavelet coefficients



(a) Geronimo (b) Donovan

Figure 5.5: Scaled multiwavelet coefficients

5.4.5 Results

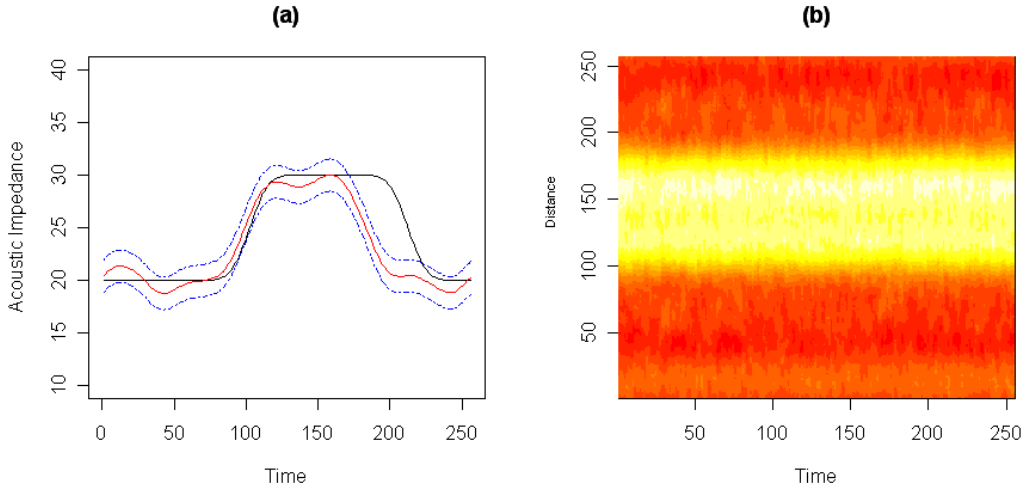
The different wavelet filters are compared based on how well they can be used in the seismic modelling examples used in Chapter 3. Here the transformed wavelet coefficients are used to build a linear regression model of the true acoustic impedance based on a known trace. The model is then ised to predict the underlying acoustic impedance distribution based on the transformed wavelet coefficients of similar recorded traces. There are two test datasets and for each, the MSE of the reconstruction using the various coefficient transform schemes. The choice of transfer function was investigated in each case and again, a running median approach was preferred and a window of 15 was used.

Original dataset

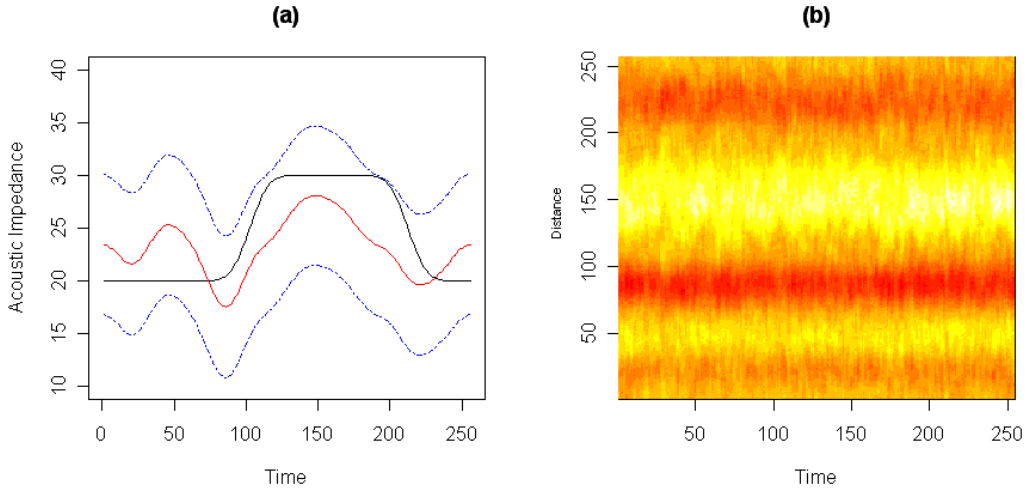
Table 5.1 shows the mean squared error of the reconstructions. This shows that by far the best performing method is a combination of the GDHM filter and splitting the coefficients. For all the other proposed methods, the reconstruction MSE is greater than in the scalar wavelet models of Chapter 3.

Alternative dataset

The modelling methods were then applied to the alternative dataset. The best performance was again from the split coefficients, the reconstruction can be seen in Figure 5.7. The MSE of the reconstruction in this case was 1.14 for the GDHM filter and 3.82 for the Donovan filter. The GDHM performance here is marginally better than the scalar wavelet neighbour model, which had an MSE of 1.16.

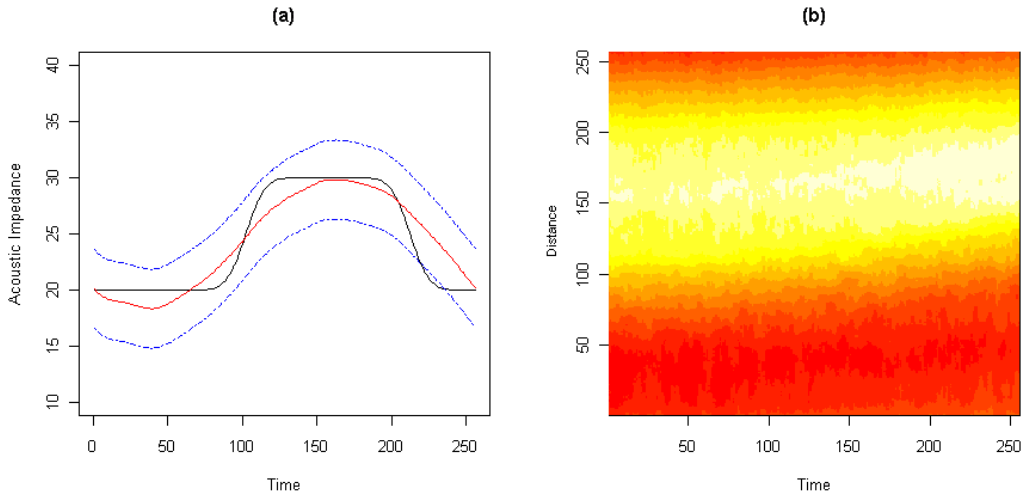


(a) Geronimo

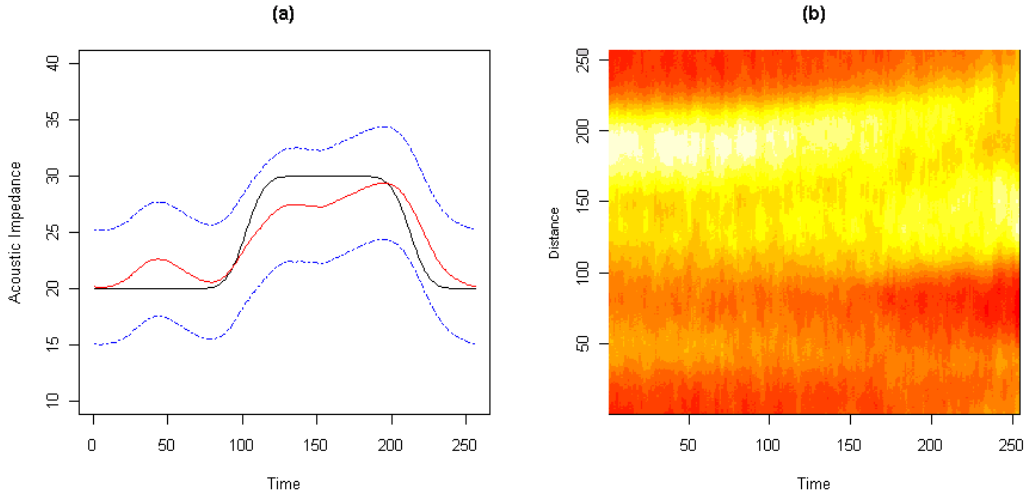


(b) Donovan

Figure 5.6: (a) Representative reconstruction and (b) Mean image reconstruction of acoustic impedance using the split coefficients



(a) Geronimo



(b) Donovan

Figure 5.7: (a) Representative reconstruction and (b) Mean image reconstruction of acoustic impedance of alternative dataset using the split coefficients

Wavelet Type	Coefficient Transform	Prediction MSE
Geronimo	Split	0.67
	Modulus	8.88
	Modulus-argument	4.05
	Scaled	9.92
Donovan	Split	9.57
	Modulus	11.31
	Modulus-argument	9.42
	Scaled	16.01

Table 5.1: Reconstruction MSE for the original dataset models.

5.5 Modelling tomographic data

5.5.1 Motivation

It was proposed in Section 5.4 that multiwavelet transforms may offer some improvement on scalar wavelets and this was in someway supported by the improvement in the reconstruction MSE in Section 5.4.5. The aim is to now explore whether using a multiwavelet transform the application setting of Chapter 4 can offer any improvement.

5.5.2 Model framework

The modelling process described in Section 4.3.5 was used to allow comparison to the scalar wavelet case. In this experiment, the bubble and churn extremes were compared to one intermediate flow regime and a multinomial model used to predict flow regimes depending upon the activity measures.

In the tomography datasets, there were 49 series observed over time. In order to allow

a multiwavelet approach to be used without the need for prefiltering, these series will be grouped into groups of 2 or 3 depending on the dimension of the filter used. In order to incorporate all of the series, and as 49 is neither a multiple of 2 nor 3, the set of series was randomly sampled to provide a set of series with the correct dimension. This should not impact on the validity of the method as each of the series were recorded from the same setup of flow regimes types.

5.5.3 Calculating activity measures

The wavelet coefficients obtained will now vary in dimension and in number depending on the filter used. The activity measures will be calculated using the different methods for incorporating the coefficients into the modelling process as described in Section 5.4.2. Let $p = ceiling(49/m)$, where m is the dimension of the filter used and $ceiling$ is a function which rounds up to the nearest integer.

Split coefficients The result of the MWT will give p vector coefficients at resolution level, $j = 1, \ldots, J$ and location k. The mean of each component of the coefficients was taken to give a vector valued activity measure. Let $\mathbf{d}^i$ be the ith component of a vector, then

$$\mathbf{a_{jk}}^i = \frac{1}{p} \sum_{x=1}^{p} \mathbf{d_{jk}}^x.$$

These are then used in the model build, meaning that there are now $p \times J$ explanatory variables.

Modulus

The modulus, r_{jk}, of each multiwavelet coefficient was found and the activity measure was taken to be the mean of these moduli at each resolution level and location over the p coefficients.

Scaled

		True Flow Regime		
		bubble	churn	intermediate
Predicted	bubble	0.85	0.12	0.19
Flow	churn	0.04	0.81	0.05
Regime	intermediate	0.08	0.07	0.76

Table 5.2: Matrix of the correct cross-classification rates between the three flow regimes using the GDHM filter (split coefficients) with added white noise standard deviation equal to 0.05.

Let the set of scaled coefficients be denoted by $\{\theta_{jk}^1, \ldots, \theta_{jk}^p\}$. Then the mean at each resolution and location of the set of scaled coefficients can be used to represent the activity measure

$$a_{jk} = \frac{1}{p} \sum_{x=1}^{p} \theta_{jk}^x.$$

5.5.4 Results

The best performance was obtained by using the split coefficient transform. The activity measures for the modulus and scaled versions of the multiwavelet coefficients were unstable and the resulting models performed poorly. For this reason, only the correct classification rates using the split coefficients using the GDHM and Donovan filter are reported in Tables 5.2 and 5.3 respectively. Figure 5.8 shows the correct classifiaction rates for the GDHM and Donovan filters. This shows that the GDHM filter performs better, yet both curves decrease steadily as the noise standard deviation increases.

		True Flow Regime		
		bubble	churn	intermediate
Predicted	bubble	0.67	0.08	0.19
Flow	churn	0.10	0.80	0.07
Regime	intermediate	0.23	0.12	0.74

Table 5.3: Matrix of the correct cross-classification rates between the three flow regimes using the Donovan filter (split coefficients) with added white noise standard deviation equal to 0.05.

5.6 Multiwavelet thresholding

5.6.1 Introduction

In Chapter 4, thresholding was shown to increase the correct classification rate when the amplitude of the noise was less than the signal present in the test data. It was therefore proposed that a multiwavelet thresholding technique may improve the classification results.

Downie and Silverman (1998) introduced the concept of thresholding in a multiwavelet setting. They showed that for an $L-$dimensional wavelet basis, a transform of the coefficients of pure noise would have a χ^2_L distribution. A threshold for the transformed coefficients can then be determined from this distribution.

The transform of the multiwavelet coefficient used in the thresholding rule is $\theta^2_{jk} = D^T_{jk} V_j^{-1} D_{jk}$, where D_{jk} are L-vector coefficients. As mentioned above, it has been shown that when considering the multiwavelet transform of pure noise, $\theta^2_{jk} \sim \chi^2_L$. Downie and Silverman (1998) give a multivariate universal threshold, which is used later. This has the form:

$$\lambda^2_n = 2\log(n) + (L-2)\log\log(n).$$

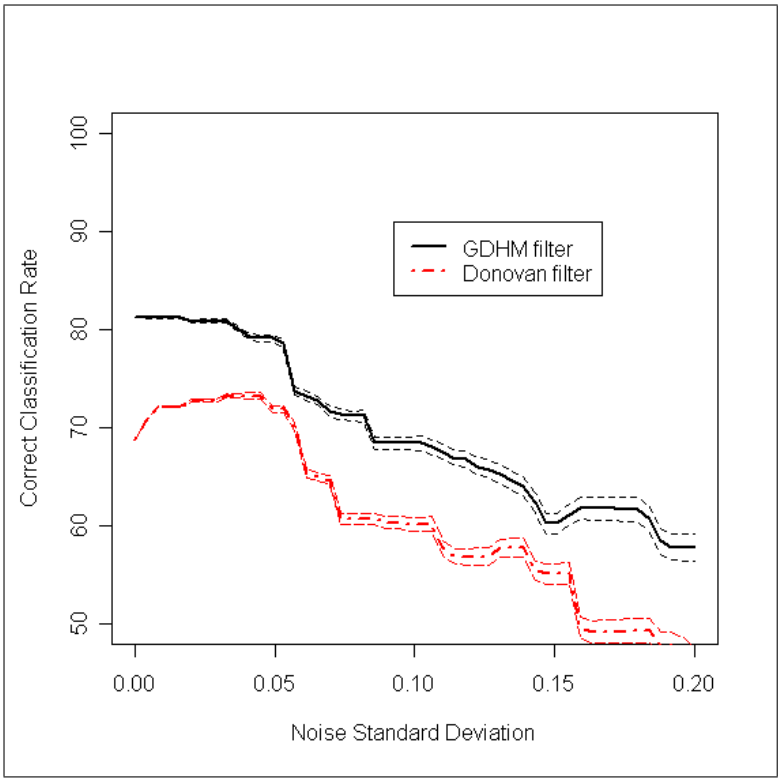


Figure 5.8: Correct classification rates of models built using the split coefficients.

The estimation of the covariance matrix V_j will then determine the orientation of the ellipse shown in Figure 5.9.

Hard multiwavelet thresholding

Looking at Figure 5.9, hard thresholding would mean that all coefficients outside of the ellipse are maintained, whilst all those inside are shrunk to zero.

$$\mathbf{D_{jk}}^* = \begin{cases} 0 & \text{if } \theta_{jk} < \lambda \\ \mathbf{D_{jk}} & \text{if } \theta_{jk} \geq \lambda \end{cases}$$

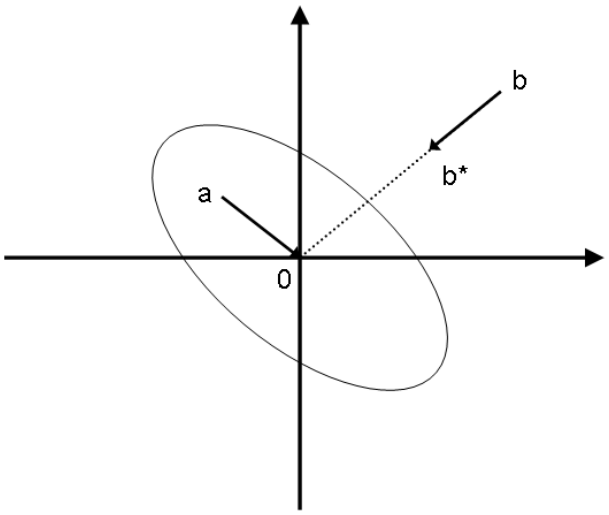


Figure 5.9: Diagram to show how a two dimensional wavelet coefficient would be transformed by thresholding.

Soft multiwavelet thresholding

For soft thresholding, all coefficients inside the ellipse are still shrunk to zero. However, in this case all the coefficients outside the ellipse have their modulus reduced by an amount equal to the radius of the ellipse along the coefficient vector. In the two dimensional case, this would mean that the point b would be mapped to the point b^*.

$$\mathbf{D_{jk}}^* = \begin{cases} 0 & \text{if } \theta_{jk} < \lambda \\ \mathbf{D_{jk}} \frac{(\theta_{jk}^2 - \lambda^2)}{\theta_{jk}^2} & \text{if } \theta_{jk} \geq \lambda \end{cases}$$

5.6.2 Application to tomographic classification

The multinomial modelling process described earlier in this Chapter was used again to allow comparison between cases. The GDHM filter was used this time, since this performed better in the non-thresholded simulations. The thresholding methods were applied to the multiwavelet coefficients obtained from the MWT, before calculating the

		True Flow Regime		
		bubble	churn	intermediate
Predicted	bubble	0.38	0.24	0.30
Flow	churn	0.41	0.31	0.43
Regime	intermediate	0.21	0.45	0.27

Table 5.4: Correct cross-classification rates using the hard thresholded split coefficient and the GDHM filter with added white noise standard deviation equal to 0.05.

		True Flow Regime		
		bubble	churn	intermediate
Predicted	bubble	0.22	0.26	0.33
Flow	churn	0.21	0.35	0.39
Regime	intermediate	0.39	0.12	0.28

Table 5.5: Correct cross-classification rates between the flow regimes using the soft thresholded split coefficients and the GDHM filter with added white noise standard deviation equal to 0.05.

activity measures. The correct classification rates of the different transfer functions show that thresholding has less affect on the prediction rates than in the scalar case.

The reason for this poor performance could be related to the way in which the multiple series are used. Downie and Silverman (1998) noted that multiwavelet thresholding gives poor results when the identity prefilter is used. The problem is that if a signal component is present at a particular time-frequency it may be represented in one element of the coefficient vector . This means that noise in the other elements will not be removed since the thresholding technique takes account of the whole vector. As the series considered here are highly correlated with each other, a similar issue may arise.

The Thresholded model performed much worse than the non-thresholded version in terms

of being able to correctly classify the flow regimes. This can be seen by comparing Tables 5.4 and 5.5 to the Table 5.2 in the previous section.

5.7 Conclusions

It was hoped that the multidimensional frequency-time decomposition offered by the multiwavelet transform would improve on the frequency modelling methods used in Chapters 3 and 4. However, the application to these two modelling scenarios have resulted in varying degrees of success.

For the seismic modelling, this new approach was able to reduce the mean squared error of the acoustic impedance reconstruction by naturally incorporating neighbouring traces into the model. The MSE was reduced to 0.67 from a minimum of 1.11 using scalar wavelets. The way in which the multiwavelet coefficients are incorporated into the model was an important factor in determining the success. Only the split coefficient method, using the GDHM filter resulted in an improvement.

In the tomographic modelling attempt, all methods performed less successfully than the scalar counterpart. Again, the GDHM filter performed better than the Donovan filter. This could be due to the symmetric/antisymmetric property, which the Donovan filter does not possess. The way in which the activity measures were calculated meant that some of the model were unstable, classifying no better than random.

This instability with regard to the way the multiwavelet coefficients are transformed would be interesting to investigate if further time were available.

In addition, the implementation of higher dimensional multiwavelet filters could prove to be more successful on higher dimensional datasets, as in the tomographic example.

Chapter 6

Conclusions and further work

The main theme of this thesis has been to consider the wavelet transform as a method of exploiting frequency characteristics that occur in some naturally occurring processes. In particular, we have looked at situations were the processes involve multiple observed series and have attempted to combine these observations in the process of building predictive models.

In Chapter 3, we considered the modelling of a seismic dataset. Using a non-decimated wavelet transformation to exploit the frequency characteristics of the data proved useful in building a predictive model. It was found that transforming the wavelet coefficients using a *transfer function* which exploited these frequency characteristics improved the reconstructive ability of the models.

It would be interesting to investigate the power of these methods on a real seismic study. However, as an initial known truth is required to build models in the way proposed, such datasets are hard to acquire.

In Chapter 4 we investigated how the localised time-frequency representation achieved by using the wavelet coefficients could be used to classify tomographic flow regimes. The introduction of the transfer function in the application allowed multiple series to be

combined to explain the frequency characteristics across all of the series. This proved a useful measure in predicting flow regimes using logistic and multinomial regression methods.

Chapter 5 looked at the application of multiwavelet bases to frequency modelling. To achieve this the non-decimated version of the multiwavelet transform was used with two and three dimensional bases. This allowed multiple series to be combined during the wavelet decomposition. This technique was used in the same applications as in Chapters 3 and 4 to see if this improved on the scalar wavelet efforts. In the seismic case, a two dimensional filter was able to reduce the reconstruction error over scalar wavelets. When using the method on tomographic data the best results were obtained from splitting the coefficients, however this offered no improvement over the scalar wavelet model.

Further investigation could be done into more dimensions of the multiwavelet approach. In the seismic application, the dimension of filter was chosen so as to include neighbouring traces in the modelling process. However in the tomography application, the bases were chosen due to their dimension. It would be of interest to extend the 'bank' of multiwavelet filters, as in the tomography setting a seven dimensional filter would have been of interest, due to the seven sensors in the experimental setup.

In this thesis, primarily the non-decimated wavelet and non-decimated multiwavelet transforms have been considered. There are several other transforms, such as the wavelet packet transform, the locally stationary wavelet transform and the complex wavelet transform. These could be investigated for modelling frequency characteristics in a similar manner.

Another important application of wavelets is in denoising. We have not considered this here as it was the underlying frequency property of the noise that we have aimed to extract. Another angle of modelling data with these characteristics might be to extract the noise using denoising methods and then model using the denoised dataset alongside the extracted data.

Bibliography

F. Abramovich, T. Sapatinas, and B. W. Silverman. Wavelet thresholding via a bayesian approach. *Journal of the Royal Statistical Society, Series B*, 60:725–749, 1998.

F. Abramovich, T. C. Bailey, and T. Sapatinas. Wavelet analysis and its statistical applications. *Journal of the Royal Statistical Society, Series D*, 48:1–30, 2000.

R. G. Aykroyd, J. G. B. Haigh, and S. Zimeras. Unexpected spatial patterns in exponential family auto models. *Graphical models and image processing*, 58:452–463, 1996.

S. Barber and G. P. Nason. Real nonparametric regression using complex wavelets. *Journal of the Royal Statistical Society, Series B*, 66:927–939, 2004.

S. Barber, G. P. Nason, and B. W. Silverman. Posterior probability intervals for wavelet thresholding. *Journal of the Royal Statisitcal Society, Series B*, 64:189–205, 2002.

E. P. G. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden Day, San Francisco, 1976.

S. Butterworth. On the theory of filter amplifiers. *Wireless Engineering*, 7:536–541, 1930.

A. Chakraborty and D. Okaya. Frequency-time decomposition of seismic data using wavelet-based methods. *Geophysics*, 60:1906–1916, 1995.

K. S. Cheng, D. Isaacson, J. C. Newell, and D. G. Gisser. Electrode models for electric current computed tomography. *IEEE Transactions on Biomedical Engineering*, 36: 918–924, 1989.

H. A. Chipman and L. J. Wolfson. Prior elicitation in the wavelet domain. In P. Müller and B. Vidakovic, editors, *Lecture Notes in Statistics: Bayesian Inference in Wavelet Based Models*, volume 141, New York, 1999. Springer-Verlag.

A. Cohen, I. Daubechies, and P. Vial. Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, 1:54–81, 1993.

T. F. Cox and M. A. A. Cox. *Multidimensional Scaling*. Chapman & Hall, 2001.

I. Daubechies. *Ten Lectures on Wavelets (C B M S - N S F Regional Conference Series in Applied Mathematics)*. SIAM, Philadelphia, 1992.

I. Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41:909–966, 1988.

D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81:425–455, 1994.

D.L. Donoho and I.M Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.*, 90:1200–1224, 1995.

G. Donovan, J. S. Geronimo, and D. P. Hardin. Intertwining multiresolution analyses and the construction of piecewise-polynomial wavelets. *SIAM Journal of Mathematical Analysis*, 27:1791–1815, 1996.

T. R. Downie and B. W. Silverman. The discrete multiple wavelet transform and thresholding methods. *IEEE Transactions on Signal Processing*, 46:2558–2561, 1998.

J. J. Faraway. *Extending the Linear Model with R*. Chapman & Hall, Boca Raton, 2005.

P. Fryzlewicz, V. Delouille, and G. P. Nason. GOES-8 x-ray sensor variance stabilization using the multiscale data-driven haar-fisz transform. *Journal of the Royal Statistical Society, Series C*, 56:99–116, 2007.

J. Geronimo, G.C. Donovan, D.P. Hardin, and P. Massopust. Construction of orthogonal wavelets using fractal interpolation functions. *SIAM J. Math. Anal*, 27:1158–1192, 1996.

M. Ghazel, G. H. Freeman, E. R. Vrscay, and R. K. Ward. Wavelet image denoising using localised thresholding operators. In *Lecture Notes in Computer Science: Image Analysis and Recognition*, volume 3656, Berlin, 2005. Springer.

R. Gnanadesikan. *Methods for statistical data analysis of multivariate observations (2nd Edition)*. Wiley, 1997.

P. J. Green and X. L. Han. Metropolis methods, gaussian proposals, and antithetic variables. *Lecure notes in Statistics*, 74:142–164, 1992.

B. Hoyle. Industrial process tomography progress. In T. Dyakowski, A. Plaskowski, D. Sankowski, Ł. Mazurkiewicz, and J. Nowakowski, editors, *Proc. 3rd International Symposium on Process Tomography in Poland — Łódź*, page 9, 2004.

P. J. Huber. *Robust Statistics*. John Wiley & Sons, New York, 1981.

U. Z. Ijaz, J. H. Kim, M. C. Kim, S. Kim, J. W. Park, and K. Y. Kim. Nondestructive dynamic process monitoring using electrical capacitance tomography. *Key Engineering Materials*, 321:1671–1674, 2006.

F. Keinert. *Wavelets and Multiwavelets*. Chapman and Hall / CRC Press, Boca Raton, Fl., 2004.

T. W. Körner. *Fourier Analysis*. Cambridge University Press, Cambridge, 1996.

S. G. Mallat. Multiresolution approximations and wavelet orthonormal bases of $l^2(\mathbb{R})$. *Transactions of the American Mathematical Society*, 315:69–87, 1989.

B.F.G. Manly. *Multivariate Statistical Methods: A Primer*. Chapman & Hall/CRC, Boca Raton, 2005.

K. Melkemi and Z. Mokhtari. Wavelet bases of the interval: A new approach. *International Journal of Mathematical Analysis*, 1:1019–1030, 2007.

Y. Meyer. *Wavelets and Operators*. Cambridge University Press, 1992.

Y. Meyer. Ondelettes sur l'intervalle. *Rivista Mathematica Lebroamericana*, 7:115–133, 1991.

G. P. Nason. *WaveThresh4 Software*. Department of Mathematics, University of Bristol, Bristol, UK, 2005. URL `http://www.stats.bris.ac.uk/~wavethresh`.

G. P. Nason and B. W. Silverman. The discrete wavelet transform in s. *Journal of Computational and Graphical Statistics*, 3:163–191, 1994.

G.P. Nason and B.W. Silverman. The stationary wavelet transform and some statistical applications. *Lecture Notes in Statistics*, 103:281–300, 1995.

A. V. Oppenheim and R. W. Schafer. *Discrete time signal processing (2nd Edition)*. Prentice Hall, Englewood Cliffs, NJ, 1999.

Y. Peijie, Y. Xingyao, and Z. Guangzhi. Seismic data analysis based on fuzzy clustering. In *8th International Conference on Signal Processing*, volume 4, pages 16–20, 2006.

D. B. Percival and A. T. Walden. *Wavelet methods for time series analysis*. Cambridge University Press, Cambridge, 2000.

M. B. Priestley. *Spectral Analysis and Time Series*. Academic Pres Inc., London, 1981.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2007. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

Enders A. Robinson and Sven Treitel. *Geophysical Signal Analysis*. Prentice-Hall, New Jersey, 1980.

StatSoft, Inc. *Electronic Statistics Textbook*. StatSoft, Tulsa, 2007. URL `http://www.statsoft.com/textbook/stathome.html`.

C. Stein. Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, 9:1135–1151, 1981.

S. Stein and M. Wysession. *An introduction to seismology, earthquakes and earth structure*. Blackwell, 2002.

E. H. Stitt and K. James. Process tomography and particle tracking: Research tool and commercial diagnostic tool. In *Proc. 3rd World Congress on Industrial Process Tomography, Banff, Canada*, pages 2–10, 2003.

G. Strang and V. Strela. Short wavelets and matrix dilation equations. *IEEE Transactions on Signal Processing*, 43:108–115, 1995.

V. Strela, P. N. Heller, G. Strang, P. Topiwala, and C. Heil. The application of multiwavelet filter banks to image processing. *IEEE Transactions on Image Processing*, 8:548–563, 1999.

D. Van der Ville, T. Blu, and M. Unser. Surfing the brain — an overiew of wavelet-based techniques for fMRI data analysis. *IEEE Transactions on Signal Processing*, 54: 1256–1267, 2006.

M. Vauhkonen, W. R. B. Lionheart, L. M. Heikkinen, P. J. Vauhkonen, and J. P. Kaipio. A MATLAB package for the EIDORS project to reconstruct two-dimensional EIT images. *Physiological Measurement*, 22:107–111, 2001.

W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002.

B. Vidakovic. *Statistical Modeling by Wavelets*. John Wiley & Sons, New York, 1999.

R. M. West, R. G. Aykroyd, S. Meng, and R. A. Williams. Spatial-temporal modelling for EIT of a mixing process. In *Proc. 3rd World Congress on Industrial Process Tomography, Banff, Canada*, pages 226–232, 2003.

R. M. West, R. G. Aykroyd, S. Meng, and R. A. Williams. Markov chain monte carlo techniques and spatial-temporal modelling for medical EIT. *Physiological Measurements*, 25:181–194, 2004.

A. B. Williams and F. J. Taylor. *Electronic filter design handbook (3rd Edition)*. McGraw Hill, 1995.

R. A. Williams and C. Xie. Tomographic techniques for characterising particulate processes. *Particle & Particle Systems Characterization*, 10:252–261, 1993.

X.H.Wang, R. S. H. Istepanian, and Y. H. Song. Application of wavelet modulus maxima for microarray spot recognition. *IEEE Transactions on NanoBioScience*, 2:190–192, 2003.

A. Zerva and O. Zhang. Correlation patterns in characteristics of spatially variable seismic ground motions. *Earthquake engineering and structural dynamics*, 26:19–39, 1997.